

DEX-0113

BN

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number  
WO 02/079398 A2

(51) International Patent Classification <sup>7</sup> :	C12N	60/314,018	21 August 2001 (21.08.2001)	US
		60/315,227	27 August 2001 (27.08.2001)	US
(21) International Application Number:	PCT/US02/07355	60/318,403	10 September 2001 (10.09.2001)	US
		60/318,510	10 September 2001 (10.09.2001)	US
(22) International Filing Date:	8 March 2002 (08.03.2002)	60/322,360	14 September 2001 (14.09.2001)	US
		60/322,296	14 September 2001 (14.09.2001)	US
(25) Filing Language:	English	60/325,378	27 September 2001 (27.09.2001)	US
		60/332,486	9 November 2001 (09.11.2001)	US
(26) Publication Language:	English	60/345,399	9 November 2001 (09.11.2001)	US
		10/094,886	7 March 2002 (07.03.2002)	US

## (30) Priority Data:

60/274,194	8 March 2001 (08.03.2001)	US
60/274,281	8 March 2001 (08.03.2001)	US
60/274,322	8 March 2001 (08.03.2001)	US
60/274,849	9 March 2001 (09.03.2001)	US
60/275,578	13 March 2001 (13.03.2001)	US
60/275,579	13 March 2001 (13.03.2001)	US
60/275,601	13 March 2001 (13.03.2001)	US
60/276,000	14 March 2001 (14.03.2001)	US
60/276,776	16 March 2001 (16.03.2001)	US
60/276,994	19 March 2001 (19.03.2001)	US
60/277,239	20 March 2001 (20.03.2001)	US
60/277,327	20 March 2001 (20.03.2001)	US
60/277,338	20 March 2001 (20.03.2001)	US
60/277,791	21 March 2001 (21.03.2001)	US
60/277,833	22 March 2001 (22.03.2001)	US
60/278,152	23 March 2001 (23.03.2001)	US
60/278,894	26 March 2001 (26.03.2001)	US
60/278,999	27 March 2001 (27.03.2001)	US
60/279,036	27 March 2001 (27.03.2001)	US
60/280,233	30 March 2001 (30.03.2001)	US
60/280,802	2 April 2001 (02.04.2001)	US
60/288,052	2 May 2001 (02.05.2001)	US
60/288,066	2 May 2001 (02.05.2001)	US
60/288,228	2 May 2001 (02.05.2001)	US
60/291,766	17 May 2001 (17.05.2001)	US
60/296,693	7 June 2001 (07.06.2001)	US
60/296,856	8 June 2001 (08.06.2001)	US
60/303,230	5 July 2001 (05.07.2001)	US
60/303,237	5 July 2001 (05.07.2001)	US
60/310,913	8 August 2001 (08.08.2001)	US
60/311,978	13 August 2001 (13.08.2001)	US
60/312,191	14 August 2001 (14.08.2001)	US
60/312,916	16 August 2001 (16.08.2001)	US
60/313,182	17 August 2001 (17.08.2001)	US
60/313,626	20 August 2001 (20.08.2001)	US

## (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	10/094,886 (CIP)
Filed on	7 March 2002 (07.03.2002)
US	60/274,194 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,281 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,322 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,849 (CIP)
Filed on	9 March 2001 (09.03.2001)
US	60/275,235 (CIP)
Filed on	12 March 2001 (12.03.2001)
US	60/275,578 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,579 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,601 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/276,000 (CIP)
Filed on	14 March 2001 (14.03.2001)
US	60/276,776 (CIP)
Filed on	16 March 2001 (16.03.2001)
US	60/276,994 (CIP)
Filed on	19 March 2001 (19.03.2001)
US	60/277,239 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,327 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,338 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,791 (CIP)
Filed on	21 March 2001 (21.03.2001)
US	60/277,833 (CIP)
Filed on	22 March 2001 (22.03.2001)

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.



WO 02/079398 A2



- US 60/278,152 (CIP)  
 Filed on 23 March 2001 (23.03.2001)  
 US 60/278,894 (CIP)  
 Filed on 26 March 2001 (26.03.2001)  
 US 60/278,999 (CIP)  
 Filed on 27 March 2001 (27.03.2001)  
 US 60/279,036 (CIP)  
 Filed on 27 March 2001 (27.03.2001)  
 US 60/280,233 (CIP)  
 Filed on 30 March 2001 (30.03.2001)  
 US 60/280,802 (CIP)  
 Filed on 2 April 2001 (02.04.2001)  
 US 60/288,052 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/288,066 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/288,228 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/291,766 (CIP)  
 Filed on 17 May 2001 (17.05.2001)  
 US 60/296,693 (CIP)  
 Filed on 7 June 2001 (07.06.2001)  
 US 60/296,856 (CIP)  
 Filed on 8 June 2001 (08.06.2001)  
 US 60/303,230 (CIP)  
 Filed on 5 July 2001 (05.07.2001)  
 US 60/303,237 (CIP)  
 Filed on 5 July 2001 (05.07.2001)  
 US 60/310,913 (CIP)  
 Filed on 8 August 2001 (08.08.2001)  
 US 60/311,978 (CIP)  
 Filed on 13 August 2001 (13.08.2001)  
 US 60/312,191 (CIP)  
 Filed on 14 August 2001 (14.08.2001)  
 US 60/312,916 (CIP)  
 Filed on 16 August 2001 (16.08.2001)  
 US 60/313,182 (CIP)  
 Filed on 17 August 2001 (17.08.2001)  
 US 60/313,626 (CIP)  
 Filed on 20 August 2001 (20.08.2001)  
 US 60/314,018 (CIP)  
 Filed on 21 August 2001 (21.08.2001)  
 US 60/315,227 (CIP)  
 Filed on 27 August 2001 (27.08.2001)  
 US 60/318,403 (CIP)  
 Filed on 10 September 2001 (10.09.2001)  
 US 60/318,510 (CIP)  
 Filed on 10 September 2001 (10.09.2001)  
 US 60/322,296 (CIP)  
 Filed on 14 September 2001 (14.09.2001)  
 US 60/322,360 (CIP)  
 Filed on 14 September 2001 (14.09.2001)  
 US 60/325,378 (CIP)  
 Filed on 27 September 2001 (27.09.2001)  
 US 60/332,486 (CIP)  
 Filed on 9 November 2001 (09.11.2001)  
 US 60/345,399 (CIP)  
 Filed on 9 November 2001 (09.11.2001)
- (71) **Applicant (for all designated States except US):** CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **KEKUDA, Ramesh** [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 (US). **TCHERNEV, Velizar, T.** [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US). **LIU, Xiaohong** [CN/US]; 96 Montoya Circle, Branford, CT 06405 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). **LI, Li** [CN/US]; 56 Jerimoth Road, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **MALYANKAR, Uriel, M.** [IN/US]; 229 Branford Road, # 330, Branford, CT 06405 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **CASMAN, Stacie, J.** [US/US]; 17 Peck Street, North Haven, CT 06473 (US). **PENA, Carol, E., A.** [US/US]; 604 Orange Street, #2, New Haven, CT 06511 (US). **GANGOLLI, Esha, A.** [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). **GUSEV, Vladimir** [UA/US]; 1209 Durham Road, Madison, CT 06443 (US). **SMITHSON, Glenda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **GERLACH, Valerie** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **POCHART, Pascale, F-J.** [FR/US]; 99 Country Way, Madison, CT 06443 (US). **FERNANDES, Elma, R.** [IN/US]; 77 Florence Road #2B, Branford, CT 06405 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **SPADERNA, Steven, K.** [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). **LAROCHELLE, William, J.** [US/US]; 15 Devonshire Lane, Madison, CT 06443 (US). **ZHONG, Mei** [CA/US]; Apartment 1B, 45 Harrison Avenue, Branford, CT 06405 (US). **KHRAMTSOV, Nikolai, V.** [RU/US]; 38 Montoya Circle, Branford, CT 06405 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **HERRMANN, John, L.** [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US).
- (74) **Agent:** **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P., C., One Financial Center, Boston, MA 02111 (US).
- (81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

[Continued on next page]





MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

- (84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## **THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE**

### **FIELD OF THE INVENTION**

5

The present invention relates to novel polypeptides having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

### **BACKGROUND OF THE INVENTION**

15 Eukaryotic cells are characterized by biochemical and physiological processes, which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates or, more particularly, organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways include constituted of extracellular signaling proteins, cellular receptors that bind the signaling proteins and signal transducing components located within the cells.

20 Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, such as two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine

effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example, induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected of suffering from a condition brought on by diminished or suppressed levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest.

### SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86. The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer

between 1 and 86, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 86. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 wherein said therapeutic is the polypeptide selected from this group.

In another embodiment, the invention comprises a method for determining the presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 in a first mammalian subject, the method involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first

subject as compared to the control sample indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, the method including introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal. The promoter may or may not be the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, the method including introducing a cell sample

expressing the polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, the method including administering the polypeptide to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86; a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86; a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and the complement of any of the nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence

selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule  
5 including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

10 In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ  
15 ID NOS: 2n-1, wherein n is an integer between 1 and 86.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein the nucleic acid molecule  
20 comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86; a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86 is changed from that selected from the group consisting of the chosen  
25 sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86; and a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ  
30 ID NO: 2n-1, wherein n is an integer between 1 and 86 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID

NO: 2n, wherein n is an integer between 1 and 86, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86, or a complement of the nucleotide sequence.

5           In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen  
10   nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

          In another embodiment, the invention includes a vector involving the nucleic acid  
15   molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

20           In another embodiment, the invention involves a method for determining the presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86 in a sample, the method including providing the sample; introducing the sample to a probe that  
25   binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

          In another embodiment, the invention involves a method for determining the presence  
30   of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86 in a first mammalian subject, the method including measuring the amount of the nucleic acid in a sample from the first mammalian



subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or

5 predisposition to the disease.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are

10 described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following

15 detailed description and claims.

### DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides,

20 antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table 1

provides a summary of the NOVX nucleic acids and their encoded polypeptides.

25

**TABLE 1. Sequences and Corresponding SEQ ID Numbers**

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	CG58548-01	1	2	Neurexophilin 1 Precursor-like
1b	174307940	3	4	Neurexophilin 1 Precursor-like
1c	CG58548-02	5	6	Neurexophilin 1 Precursor-like
1d	CG58548-03	7	8	Neurexophilin 1 Precursor-like
2a	CG58542-01	9	10	Neurophilin-like
2b	169679583	11	12	Neurophilin-like
2c	169679634	13	14	Neurophilin-like
3a	CG58540-01	15	16	Cytoplasmic-Antiproteinase 3-like

4a	CG56340-03	17	18	Interferon-like
4b	174308150	19	20	Interferon-like
5a	CG58514-01	21	22	Leprecan-like
6a	CG57887-01	23	24	Tumor suppressor-like
7a	CG57885-01	25	26	Procholoecystokinin Precursor-like
8a	CG57865-01	27	28	Secreted protein-like
8b	171651532	29	30	Secreted protein-like
9a	CG54503-03	31	32	Gliacolin-like
10a	CG58600-01	33	34	Olfactomedin-like
11a	CG57572-01	35	36	CMP-N-Acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase like
12a	CG57518-01	37	38	Neural cell adhesion protein Big-2 precursor-like
12b	170108372	39	40	Neural cell adhesion protein Big-2 precursor-like
12c	170108393	41	42	Neural cell adhesion protein Big-2 precursor-like
12d	170343246	43	44	Neural cell adhesion protein Big-2 precursor-like
12e	170343692	45	46	Neural cell adhesion protein Big-2 precursor-like
12f	170684238	47	48	Neural cell adhesion protein Big-2 precursor-like
12g	170534177	49	50	Neural cell adhesion protein Big-2 precursor-like
13a	CG57409-03	51	52	Neural cell adhesion protein Big-2 precursor-like
13b	CG57409-05	53	54	MAM and Ig domain-containing protein-like
13c	CG57409-06	55	56	MAM and Ig domain-containing protein
14a	CG59262-01	57	58	Calcium binding protein S100P-like
15a	CG58635-01	59	60	S-100-like
15b	CG58635-02	61	62	Secretory carrier membrane protein-like
15c	CG58635-03	63	64	Secretory carrier membrane protein-like
16a	CG59209-01	65	66	CG3714-like
16b	174308417	67	68	CG3714-like
16c	174308429	69	70	CG3714-like
17a	CG59368-01	71	72	Preoptic regulatory factor-2-like
18a	CG58628-01	73	74	Adipophilin-like
18b	174228350	75	76	Adipophilin-like
18c	174228354	77	78	Adipophilin-like
18d	188888733	79	80	Adipophilin-like
19a	CG59342-01	81	82	FIS-like
20a	CG59486-01	83	84	Zn finger protein-like
21a	CG59446-01	85	86	Neurotransmission associated protein-like
21b	174308261	87	88	Neurotransmission associated protein-like
21c	174308266	89	90	Neurotransmission associated protein-like
21d	174308278	91	92	Neurotransmission associated protein-like
21e	174308283	93	94	Neurotransmission associated protein-like

21f	174308287	95	96	Neurotransmission associated protein-like
21g	174308293	97	98	Neurotransmission associated protein-like
21h	174308301	99	100	Neurotransmission associated protein-like
21i	174308311	101	102	Neurotransmission associated protein-like
21j	174308315	103	104	Neurotransmission associated protein-like
21k	174308321	105	106	Neurotransmission associated protein-like
21l	174308327	107	108	Neurotransmission associated protein-like
21m	174308337	109	110	Neurotransmission associated protein-like
21n	CG59446-02	111	112	Neurotransmission associated protein-like
22a	CG59375-01	113	114	Drebrin -like
23a	CG59321-01	115	116	UNC5H2-like
23b	CG59321-02	117	118	UNC5H2-like
24a	CG59591-01	119	120	Trypsin inhibitor-like
25a	CG59588-01	121	122	ISLR precursor-like
26a	CG59584-01	123	124	Ovostatin precursor-like
26b	CG59584-02	125	126	Ovostatin precursor-like
27a	CG59417-01	127	128	Chymotrypsin precursor-like
28a	CG59415-01	129	130	Laminin type EGF-like
28b	191815704	131	132	Laminin type EGF-like
28c	191815724	133	134	Laminin type EGF-like
28d	CG59415-02	135	136	Laminin type EGF-like
29a	CG59297-01	137	138	Polycystic kidney disease 1 Protein-like
30a	CG59264-01	139	140	Polycystic kidney disease 2 Protein-like
31a	CG59623-01	141	142	Slit-like
32a	CG59247-01	143	144	Protein-tyrosine sulfotransferase-like
33a	CG59430-01	145	146	Serine Protease inhibitor-like
34a	CG59305-01	147	148	Fibronectin type III-like
34b	CG59305-02	149	150	Fibronectin type III-like
35a	CG59547-01	151	152	Adipophilin-like
36a	CG58508-01	153	154	Small inducible cytokine A4 precursor -like
36b	CG58508-02	155	156	Small inducible cytokine A4 precursor -like
36c	170072532	157	158	Small inducible cytokine A4 precursor -like
36d	170072551	159	160	Small inducible cytokine A4 precursor -like
36e	170072555	161	162	Small inducible cytokine A4 precursor -like
36f	CG58508-03	163	164	Small inducible cytokine A4 precursor -like
37a	CG59819-01	165	166	Latent transforming growth factor-like
37b	CG59819-02	167	168	Latent transforming growth factor-like
37c	CG59819-03	169	170	Latent transforming growth factor-like

38a	CG59685-01	171	172	Thrombospondin-like
38b	175070296	173	174	Thrombospondin-like
38c	175070324	175	176	Thrombospondin-like
39a	CG57167-01	177	178	Urokinase plasminogen activator surface receptor-like
40a	CG59841-01	179	180	Aggrin precursor-like
41a	CG59895-01	181	182	Major urinary protein 4 precursor-like
41b	CG59895-02	183	184	Major urinary protein 4 precursor-like
42a	CG59889-01	185	186	KIAA1199-like
42b	CG59889-02	187	188	KIAA1199-like
42c	CG59889-04	189	190	KIAA1199-like
43a	CG59512-02	191	192	Small inducible cytokine A3-like
43b	CG59512-01	193	194	Small inducible cytokine A3-like
44a	CG56801-02	195	196	Thrombomodulin-like

Table 1 indicates homology of NOVX nucleic acids to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table 1 will be useful in

5 therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table 1.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of

10 domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table 1, the NOVX polypeptides of the present invention show homology to, and contain

15 domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Examples 1-44.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small

20 molecules that modulate or inhibit diseases associated with the protein families listed in Table 1.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example 47. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds

according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, e.g. a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the  
5 invention are disclosed herein.

### NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of  
10 domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological  
15 conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential  
20 diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule  
25 drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino  
30 acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more

than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1

and 86; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

### NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide, precursor form, or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (host cell) in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from

a post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

5           The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), and 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and  
10 much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

          The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid which is separated from other nucleic acid molecules which are present in the natural source of the  
15 nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, or less of nucleotide sequences which naturally flank  
20 the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, culture medium, or of chemical precursors or other chemicals.

          A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the  
25 nucleotide sequence SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 86, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and  
30 cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993).



A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore,  
5 oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar  
10 or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, or a complement  
15 thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-  
20 active portion of A NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and  
25 86, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van  
30 der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

“Fragments” provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and are at most some portion less than a full length sequence. Fragments may be  
5 derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction  
10 of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

“Derivatives” are nucleic acid sequences or amino acid sequences formed from the  
15 native compounds either directly, by modification, or by partial substitution. “Analogous” are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound, e.g. they differ from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are  
20 nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or  
25 proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins of the invention under  
30 stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A “homologous nucleic acid sequence” or “homologous amino acid sequence,” or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those

sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for A NOVX polypeptide of  
5 species other than humans, including, but not limited to vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding a human NOVX protein.  
10 Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX  
15 nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with  
20 or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning  
25 NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an  
30 integer between 1 and 86; or an anti-sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86; or of a naturally occurring mutant of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various

embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express A NOVX protein, such as by measuring a level of A NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of A NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, that encodes a polypeptide having A NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

#### **NOVX Nucleic Acid and Polypeptide Variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding A NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the

result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:2n-1, wherein  
5 n is an integer between 1 and 86, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

10 Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, 2000 or more nucleotides in length. In yet another  
15 embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other  
20 than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no  
25 other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength, pH and nucleic acid concentration) at  
30 which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess at T<sub>m</sub>, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at

pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

- 5 Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are
- 10 hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65 °C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, corresponds to a
- 15 naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

- In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an
- 20 integer between 1 and 86, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are
- 25 well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

- In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer between 1
- 30 and 86, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40 °C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH

7.4), 5 mM EDTA, and 0.1% SDS at 50 °C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

### Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of the NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 86. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2n, wherein n is an integer between 1 and 86. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; more preferably at least about 70% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; still more preferably at least about 80% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; even more preferably at least about 90% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86; and most preferably at least about 95% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

An isolated nucleic acid molecule encoding A NOVX protein homologous to the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 86, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of A NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or



biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and A NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to  
5 regulate a specific biological function (e.g., regulation of insulin release).

#### Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the  
10 nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that  
15 comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of A NOVX protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 86, or antisense nucleic acids complementary to A NOVX nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1  
20 and 86, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding A NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons, which are translated into amino acid residues. In another embodiment, the antisense nucleic acid  
25 molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences, which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein,  
30 antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation

start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) 5 can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic 10 acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 15 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 20 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

25 The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding A NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an 30 antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified

such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector  
5 constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. A  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the  
10 strands run parallel to each other. See, e.g., Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (See, e.g., Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (See, e.g., Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330).

#### 15 **Ribozymes and PNA Moieties**

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in  
20 therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in  
25 Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of A NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be  
30 constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, e.g., U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be

used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g., Bartel et al., (1993) Science 261:1411-1418.*

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g., the NOVX promoter and/or enhancers*) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g., Helene, 1991. Anticancer Drug Des. 6: 569-84; Helene, et al. 1992. Ann. N.Y. Acad. Sci. 660: 27-36; Maher, 1992. Bioassays 14: 807-15.*

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g., the stability, hybridization, or solubility of the molecule.* For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g., Hyrup, et al., 1996. Bioorg Med Chem 4: 5-23.* As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g., DNA mimics*) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al., 1996. supra; Perry-O'Keefe, et al., 1996. Proc. Natl. Acad. Sci. USA 93: 14670-14675.*

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g., inducing transcription or translation arrest or inhibiting replication.* PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S<sub>1</sub> nucleases (See, Hyrup, et al., 1996. supra); or as probes or primers for DNA sequence and hybridization (See, Hyrup, et al., 1996, supra; Perry-O'Keefe, et al., 1996. supra).*

In another embodiment, PNAs of NOVX can be modified, *e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art.* For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g., RNase H and DNA polymerases*) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking,

number of bonds between the nucleobases, and orientation (*see*, Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, et al., 1996. *supra* and Finn, et al., 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and  
5 modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See*, e.g., Mag, et al., 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See*, e.g., Finn, et al., 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA  
10 segment and a 3' PNA segment. *See*, e.g., Petersen, et al., 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see*, e.g., Letsinger, et al., 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, et al., 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see*, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see*, e.g., Krol, et al., 1988. *BioTechniques* 6:958-976) or intercalating agents (*see*, e.g., Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another  
20 molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

### NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino  
25 acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2n, wherein n is an integer between 1 and 86. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

30 In general, A NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed

by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, A NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 86) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of A NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of A NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown SEQ ID NOS:2n, wherein n is an integer between 1 and 86. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 86, and retains the functional activity of the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 86, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 86, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2n, wherein n is an integer between 1 and 86.

#### DETERMINING HOMOLOGY BETWEEN TWO OR MORE SEQUENCES

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known

in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above  
5 exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of  
10 comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*,  
15 the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a  
20 reference sequence over a comparison region.

#### CHIMERIC AND FUSION PROTEINS

The invention also provides NOVX chimeric or fusion proteins. As used herein, A NOVX "chimeric protein" or "fusion protein" comprises A NOVX polypeptide operatively-  
25 linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to A NOVX protein SEQ ID NOS:2n, wherein n is an integer between 1 and 86, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from  
30 the same or a different organism. Within A NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of A NOVX protein. In one embodiment, A NOVX fusion protein comprises at least one biologically active portion of A NOVX protein. In another embodiment, A NOVX fusion protein comprises at least two biologically active portions of A NOVX protein. In yet another embodiment, A NOVX fusion protein comprises at least three



biologically active portions of A NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

5 In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is A NOVX protein containing a  
10 heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the  
15 immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between A NOVX ligand and A NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of A NOVX cognate ligand. Inhibition  
20 of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with A  
25 NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction  
30 enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two

consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.,* a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

#### NOVX AGONISTS AND ANTAGONISTS

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.,* discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade, which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.,* truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.,* for phage display) containing the set of NOVX sequences therein. There are a variety of methods, which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the

desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well known within the art. See, e.g., Narang, 1983. *Tetrahedron* 39: 3; Itakura, et al., 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, et al., 1984. *Science* 198: 1056; Ike, et al., 1983. *Nucl. Acids Res.* 11: 477.

5

#### **POLYPEPTIDE LIBRARIES**

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of A NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of A NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S<sub>1</sub> nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, et al., 1993. *Protein Engineering* 6:327-331.

#### **NOVX Antibodies**

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that

contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F<sub>ab</sub>, F<sub>ab'</sub> and F<sub>(ab)<sub>2</sub></sub> fragments, and an F<sub>ab</sub> expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ  
5 from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or  
10 fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the  
15 full length protein, such as an amino acid sequence shown in SEQ ID NOs: 2n, wherein n is an integer between 1 and 86, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at  
20 least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will  
25 indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier  
30 transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal  
5 or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

#### 10 **Polyclonal Antibodies**

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic  
15 protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further  
20 include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents.  
25 Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide  
30 primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson

(The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

### **Monoclonal Antibodies**

5       The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus  
10       contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to  
15       elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human  
20       mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are  
25       employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which  
30       substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San

Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light

chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

### **Humanized Antibodies**

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

### **Human Antibodies**

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein.



Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison ( *Nature* 368, 812-13 (1994)); Fishwild et al., (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse<sup>TM</sup> as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively

from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

5 An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

10 A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

20 In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

## 25 **F<sub>ab</sub> Fragments and Single Chain Antibodies**

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F<sub>ab</sub> expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F<sub>ab</sub> fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F<sub>(ab)<sup>2</sup></sub> fragment produced by pepsin digestion of an antibody molecule; (ii) an F<sub>ab</sub> fragment generated by reducing the disulfide bridges of an F<sub>(ab)<sup>2</sup></sub> fragment; (iii) an F<sub>ab</sub>

fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F<sub>v</sub> fragments.

### **Bispecific Antibodies**

5        Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

      Methods for making bispecific antibodies are known in the art. Traditionally, the  
10       recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct  
15       bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

      Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion  
20       preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-  
25       transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

      According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part  
30       of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a

mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')<sub>2</sub> bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')<sub>2</sub> fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')<sub>2</sub> molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V<sub>H</sub>) connected to a light-chain variable domain (V<sub>L</sub>) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V<sub>H</sub> and V<sub>L</sub> domains of one fragment are forced to pair with the complementary V<sub>L</sub> and V<sub>H</sub> domains of another fragment, thereby forming two antigen-binding

sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

- 5 Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as
- 10 to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds
- 15 tissue factor (TF).

#### **Heteroconjugate Antibodies**

Heteroconjugate antibodies are also within the scope of the present invention.

- Heteroconjugate antibodies are composed of two covalently joined antibodies. Such
- 20 antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by
- 25 forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

#### **Effector Function Engineering**

- 30 It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and

antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody  
 5 can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

### **Immunoconjugates**

The invention also pertains to immunoconjugates comprising an antibody conjugated  
 10 to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used  
 15 include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolacca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, croton, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of  
 20 radionuclides are available for the production of radioconjugated antibodies. Examples include  $^{212}\text{Bi}$ ,  $^{131}\text{I}$ ,  $^{131}\text{In}$ ,  $^{90}\text{Y}$ , and  $^{186}\text{Re}$ .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl  
 25 adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in  
 30 Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is

administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

5           **Immunoliposomes**

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545.

10       Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody  
15       of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

20           **Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention**

Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic  
25       methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds (see below).

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such  
30       an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a

clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

#### **Antibody Therapeutics**

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount



required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

#### **Pharmaceutical Compositions of Antibodies**

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example,

hydroxymethylcellulose or gelatin-microcapsules and poly-(methacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

- 5           The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or  
10   microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and  $\gamma$  ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON  
DEPOT<sup>TM</sup> (injectable microspheres composed of lactic acid-glycolic acid copolymer and  
15   leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

#### ELISA Assay

- 20           An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F<sub>ab</sub> or F<sub>(ab)2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a  
25   detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with  
fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues,  
30   cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro*

techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

#### 15            **NOVX Recombinant Expression Vectors and Host Cells**

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the

recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

10        Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

15        One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g., 20        Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

      In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 25        933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

      Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the 30        pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

      In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are

often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, *e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a

recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

## 5           **Transgenic NOVX Animals**

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous  
10 NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in  
15 which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types  
20 or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

25           A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a  
30 non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of



NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-  
10 encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of A NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can  
15 be a human gene (*e.g.*, the cDNA of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is  
20 designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional  
25 protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The  
30 additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX

gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g., Li, et al., 1992. Cell 69: 915.*

The selected cells are then injected into a blastocyst of an animal (*e.g., a mouse*) to form aggregation chimeras. *See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND*  
5 *EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152.* A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for  
10 constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol. 2: 823-829*; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a  
15 system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 6232-6236.* Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See, O'Gorman, et al., 1991. Science 251:1351-1355.* If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing  
20 transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g., by* mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced  
25 according to the methods described in Wilmot, *et al., 1997. Nature 385: 810-813.* In brief, a cell (*e.g., a somatic cell*) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g., through the use of* electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to  
30 morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g., the somatic cell*) is isolated.

### Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs  
5 and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like,  
10 compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be  
15 used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its  
20 intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine,  
25 propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral  
30 preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration,

suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, A NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient

such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

- 5           For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

- Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be  
10 permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

- 15           The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

- In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release  
20 formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal  
25 suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

- It is especially advantageous to formulate oral or parenteral compositions in dosage  
30 unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent

on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as  
5 gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery  
10 vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

15

#### Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in A NOVX gene,  
20 and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds  
25 and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention  
30 can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

### Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of A NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g., Lam, 1997. Anticancer Drug Design 12: 145.*

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science*

249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to A NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the ability of the test compound to interact with A NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with A NOVX target molecule. As used herein, a "target molecule" is a molecule with which A NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses A NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or A NOVX



protein or polypeptide of the invention. In one embodiment, A NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with A NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with A NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising A NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting A NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the ability of the test compound to interact with A NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to A NOVX target molecule by one of the methods described

above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate A NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate  
5 can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with A NOVX protein, wherein determining the  
10 ability of the test compound to interact with A NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of A NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the  
15 membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton<sup>®</sup> X-100, Triton<sup>®</sup> X-114, Thesit<sup>®</sup>,  
20 Isotridecypoly(ethylene glycol ether)<sub>n</sub>, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of  
25 complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a  
30 fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is

incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the  
5 complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated  
10 NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule,  
15 can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target  
20 molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of  
25 NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein  
30 expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8:

5 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors,  
10 which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation  
15 domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming A NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be  
20 detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

## 25           **Detection Assays**

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii)  
30 identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

### Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called  
5 chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

10 Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only  
15 those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By  
20 using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy  
25 mapping of individual genes to specific human chromosomes. *See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924.* Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular  
30 sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one

step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually.

- 5 The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A  
10 MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

- Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more  
15 likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

- Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line  
20 through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

- Moreover, differences in the DNA sequences between individuals affected and  
25 unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome  
30 spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

### **Tissue Typing**

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

### **Predictive Medicine**

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for

prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in A NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

#### DIAGNOSTIC ASSAYS

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The



nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 86, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of

NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

10

#### PROGNOSTIC ASSAYS

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be

administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in A NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding A NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from A NOVX gene; (ii) an addition of one or more nucleotides to A NOVX gene; (iii) a substitution of one or more nucleotides of A NOVX gene, (iv) a chromosomal rearrangement of A NOVX gene; (v) an alteration in the level of a messenger RNA transcript of A NOVX gene, (vi) aberrant modification of A NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of A NOVX gene, (viii) a non-wild-type level of A NOVX protein, (ix) allelic loss of A NOVX gene, and (x) inappropriate post-translational modification of A NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in A NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to A NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated

that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification  
5 system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q $\beta$  Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

10 In an alternative embodiment, mutations in A NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in  
15 the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing  
20 hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base  
25 changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the  
30 other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and

- Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S<sub>1</sub> nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on A NOVX sequence, e.g., a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP)

may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g., Orita, et al., 1989. Proc. Natl. Acad. Sci. USA: 86: 2766; Cotton, 1993. Mutat. Res. 285: 125-144; Hayashi, 1992. Genet. Anal. Tech. Appl. 9: 73-79.*

Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured  
5 and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment,  
10 the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g., Keen, et al., 1991. Trends Genet. 7: 5.*

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient  
15 gel electrophoresis (DGGE). *See, e.g., Myers, et al., 1985. Nature 313: 495.* When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See,*  
20 *e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.*

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit  
25 hybridization only if a perfect match is found. *See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230.* Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

30 Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448*) or at the extreme 3'-terminus of one primer where,

under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments  
5 amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing  
10 pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving A NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which  
15 NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

#### PHARMACOGENOMICS

20 Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's  
25 Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may  
30 be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to

determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

- 5            Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on
- 10 the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials,
- 15 sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

- As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome Pregnancy Zone Protein Precursor enzymes CYP2D6 and CYP2C19) has
- 20 provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly
- 25 polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At
- 30 the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate



agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when  
5 treating a subject with A NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

#### MONITORING OF EFFECTS DURING CLINICAL TRIALS

10 Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be  
15 monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such  
20 clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates  
25 NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described  
30 herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of A NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

#### Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

#### DISEASE AND DISORDERS

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with

Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) 5 nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, 10 agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with 15 Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or 20 RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) 25 and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

#### PROPHYLACTIC METHODS

In one aspect, the invention provides a method for preventing, in a subject, a disease or 30 condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation

of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, A NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described  
5 herein. The prophylactic methods of the invention are further discussed in the following subsections.

### Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression  
10 or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of A NOVX protein, a peptide, A NOVX peptidomimetic, or other small molecule. In  
15 one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in*  
20 *vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of A NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or  
25 combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering A NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally  
30 downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

### **Determination of the Biological Effect of the Therapeutic**

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

### **Prophylactic and Therapeutic Uses of the Compositions of the Invention**

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which

immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

5

### Examples

#### EXAMPLE 1.

The NOV1 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 1A.

10

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	813 bp	
NOV1a, CG58548-01 DNA Sequence	CAAAA <b>C</b> AAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATT <b>T</b> TAACG AACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCACA TATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTACAGACTTTTCG TGGCAAAGAGAATGATACAGATTGGACCTGAGATATGACACCCAGAACCTTATTCT GAGCAAGACCTCTGGGACTGGCTGAGGAACTCCACAGACCTTCAAGAGCCTCGGCCCA GGGCCAAGAGAAGGCCCATTTGTTAAAAACGGGCAAGTTTAAGAAAATGTTTGGATGGGG CGATTTTCATTCCAACATCAAAACAGTGAAGCTGAACCTGTTGATAACTGGGAAAATT GTAGATCATGGCAATGGGACATTTAGTGTTTATTTCAAGGCATAATTCAACTGGTCAAG GGAATGTATCTGTGAGCTTGGTACCCCTACAAAATCGTGGAATTTGACTTGGCACA ACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCCGATTGAATATGAA AAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCTTCAAAAACCTGTT ACCAGGAGCAAACCAAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTTAAGGTGAT CTGTATTTACATTTCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCT GACTACAAC <b>T</b> ACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATGGG <u>G</u>		
	ORF Start: ATG at 17	ORF Stop: TGA at 797	
	SEQ ID NO: 2	260 aa	MW at 29905.5kD
NOV1a, CG58548-01 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKH <b>I</b> WTESKDL <b>S</b> ISRLLSQTFRGKEND TDLDLRYDTPPEYSEQDLWDWLRNSTDLQEP <b>R</b> PRAKRR <b>P</b> I <b>V</b> KTGKFKMF <b>G</b> WGD <b>F</b> HSN IKTVKLNLLITGKIVDHGNGTFSVYFRHNSTGQGNVSVSLVPPTK <b>I</b> VEFDLAQQT <b>V</b> IL AKDSKSFNCRIEY <b>K</b> VDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPFK <b>V</b> IC <b>I</b> Y <b>S</b> FYSTDYKLVQKVC <b>P</b> DYNYHSDTPYFPSG		
	SEQ ID NO: 3	771 bp	
NOV1b, 174307940 DNA Sequence	GGATCCGTAACATGTGCCAATTTAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAG GAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTAT CAGCCGACTCCTGTACAGACTTTTCGTGGCAAAGAGAATGATACAGATT <b>T</b> GGACCTG AGATATGACACCCAGAACCTTATTCTGAGCAAGACCTCTGGGACTGGCTGAGGA <b>A</b> CT CCACAGACCTTCAAGAGCCTCGGCCCAAGGGAAGGAGAG		

	TTCCCTCGGGACTCGAG		
	ORF Start: GGA at 1	ORF Stop: E at 772	
	SEQ ID NO: 4	257 aa	MW at 29326.8kD
NOV1b, 174307940 Protein Sequence	GSVTCANLTNGGKSELLKSGSSKSTLKHIIWTESSKDLSISRLLSQTRFGKENDTDLDLRYDTPEPYSEQDLWDWLRNSTDLQEP RPRAKRRPIVKTGKFKMFGWGFHNSNIKTVKLNLLITGKIVDHNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKDSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPFKVICIYISFYSTDYKLVQKVCPDYNHSDTPYFPSGLE		
	SEQ ID NO: 5	813 bp	
NOV1c, CG58548-02 DNA Sequence	CAAAACAAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATTAAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTACAGACTTTTCGTGGCAAAGAGAATGATACAGATTGGACCTGAGATATGACACCCAGAACCTTATTCTGAGCAAGACCTCTGGGACTGGCTGAGGAACCTCCACAGACCTTCAAGAGCCTCGGCCCAAGGCCAAGAGAAGGCCCATTTGTTAAACGGGCAAGTTTAAGAAAATGTTTGGATGGGCGATTTTCATTCCAACATCAAAACAGTGAAGCTGAACCTGTTGATAACTGGGAAAATTGTAGATCATGGCAATGGGACATTTAGTGTTTATTTCAAGGCATAATTCAACTGGTCAAGGGAATGTATCTGTACAGCTTGGTACCCCTACAAAATCGTGGAATTTGACTTGGCACAACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCGCATTGAATATGAAAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCCTCAAAAACCTGTTACCAGGAGCAAACCCAAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTTAAGGTGATCTGTATTTACATTTCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCTGACTACAACCTACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATGGGG		
	ORF Start: ATG at 17	ORF Stop: TGA at 797	
	SEQ ID NO: 6	260 aa	MW at 29905.5kD
NOV1c, CG58548-02 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKHIIWTESSKDLSISRLLSQTRFGKENDTDLDLRYDTPEPYSEQDLWDWLRNSTDLQEP RPRAKRRPIVKTGKFKMFGWGFHNSNIKTVKLNLLITGKIVDHNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKDSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPFKVICIYISFYSTDYKLVQKVCPDYNHSDTPYFPSG		
	SEQ ID NO: 7	627 bp	
NOV1d, CG58548-03 DNA Sequence	CAAAACAAATTAAAAGATGAAGGAATACTATATCCATGTAACATGTGCCAATTAAACGAACGGTGGAAAGTCAGAACTTCTGAAATCAGGAAGCAGCAAATCCACACTAAAGCACATATGGACAGAAAGCAGCAAAGACTTGTCTATCAGCCGACTCCTGTACAGACTTTTCGTGGCAAAGAGAATGATACAGATTGAACCTGTTGATAACTGGGAAAATTGTAGATCATGGCAATGGGACATTTAGTGTTTATTTCAAGGCATAATTCAACTGGTCAAGGGAATGTATCTGTACAGCTTGGTACCCCTACAAAATCGTGGAATTTGACTTGGCACAACAAACCGTGATTGATGCCAAAGATTCCAAGTCTTTTAATTGTCGCATTGAATATGAAAAGGTTGACAAGGCTACCAAGAACACACTCTGCAACTATGACCCCTCAAAAACCTGTTACCAGGAGCAAAACCAAGTCATGTATCCTGGCTCTGCTCCAAGCCCTTAAGGTGATCTGTATTATCATTTCCTTTTATAGTACAGATTATAAACTGGTACAGAAAGTGTGCCCTGACTACAATACCACAGTGACACACCTTACTTTCCCTCGGGATGAAGGTGAACATG		
	ORF Start: ATG at 17	ORF Stop: TGA at 614	
	SEQ ID NO: 8	199 aa	MW at 22496.1kD
NOV1d, CG58548-03 Protein Sequence	MKEYYIHVTCANLTNGGKSELLKSGSSKSTLKHIIWTESSKDLSISRLLSQTRFGKENDTDLNLLITGKIVDHNGTFSVYFRHNSTGQGNVSVSLVPPTKIVEFDLAQQTVIDAKLSKSFNCRIEYKVDKATKNTLCNYDPSKTCYQEQTQSHVSWLCSKPLKVICIYISFYSTDYKLVQKVCPDYNHSDTPYFPSG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

<b>Table 1B. Comparison of NOV1a against NOV1b through NOV1d.</b>		
<b>Protein Sequence</b>	<b>NOV1a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV1b	8..260 3..255	236/253 (93%) 236/253 (93%)
NOV1c	1..260 1..260	243/260 (93%) 243/260 (93%)
NOV1d	1..260 1..199	161/260 (61%) 164/260 (62%)

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

<b>Table 1C. Protein Sequence Properties NOV1a</b>	
PSort analysis:	0.5297 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 1D.

<b>Table 1D. Geneseq Results for NOV1a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV1a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
ABB11858	Human neurexophilin homologue, SEQ ID NO:2228 - Homo sapiens, 305 aa. [WO200157188-A2, 09-AUG-2001]	8..260 53..305	253/253 (100%) 253/253 (100%)	e-152
AAB43066	Human ORFX ORF2830 polypeptide sequence SEQ ID NO:5660 - Homo sapiens, 253 aa. [WO200058473-A2, 05-OCT-2000]	8..260 1..253	253/253 (100%) 253/253 (100%)	e-152
AAM57924	Human brain expressed single exon probe encoded protein SEQ ID NO: 30029 - Homo sapiens, 235 aa. [WO200157275-A2, 09-AUG-2001]	14..248 1..235	235/235 (100%) 235/235 (100%)	e-140
AAB28778				5e-73



	fragment encoded by gene 45 - Homo sapiens, 128 aa. [WO200055198-A1, 21-SEP-2000]	1..128	128/128 (100%)	
AAB28779	Protein fragment encoded by gene 45 - Homo sapiens, 128 aa. [WO200055198-A1, 21-SEP-2000]	104..231 1..128	127/128 (99%) 127/128 (99%)	4e-72

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P58417	Neurexophilin 1 precursor - Homo sapiens (Human), 271 aa.	8..260 19..271	253/253 (100%) 253/253 (100%)	e-151
Q61200	Neurexophilin 1 precursor - Mus musculus (Mouse), 253 aa (fragment).	8..260 1..253	253/253 (100%) 253/253 (100%)	e-151
Q63366	Neurexophilin 1 precursor (Neurophilin) - Rattus norvegicus (Rat), 271 aa.	8..260 19..271	251/253 (99%) 252/253 (99%)	e-150
O95156	Neurexophilin 2 precursor - Homo sapiens (Human), 262 aa (fragment).	72..260 74..262	153/189 (80%) 170/189 (88%)	3e-93
Q28145	Neurexophilin 2 precursor (Neurophilin) - Bos taurus (Bovine), 264 aa.	72..260 76..264	153/189 (80%) 170/189 (88%)	4e-93

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1F.

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

5

#### EXAMPLE 2.

The NOV2 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 9	796 bp	
NOV2a, CG58542-01 DNA Sequence	AGGAGGAAGATGCAACTGACTCGCTGCTGCTTCGTGTTCTCGTGGTGCAGGGTAGCCTCT ATCTGGTCATCTGTGGCCAGGATGATGGTCTCTCCCGGCTCAGAGGACCCTGAGCGTGA TGACCACGAGGGCCAGCCCCGGCCCCGGGTGCTCGGAAGCGGGGCCACATCTCACCT AAGTCCCGCCCCATGGCCAATTCACCTCTCTAGGGCTGCTGGCCCCGCCTGGGGAGG CTTGGGGCATTCTTGGGCAGCCCCCAACCGCCGAACCAAGCCCCCACCCTCAGC CAAGGTGAAGAAAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCC CTGAACCTGCTCGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCC ACTTCCAACAATGCCACAGGCCAGGGAACATCTCCATCAGCCTCGTGCCCCCAG TAAAGCTGTAGAGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAA ATCTTCAACTGCCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGGCGACCTCGCTTT GCACCCACGACCCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTG GAGCTGCTCCAGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCAGGAC TATCGGCTGGTCCAGAAGGTGTGCCAGATTACAACATACCATAGTGATACCCCTACT ACCCATCTGGGTGACCCGGGGCAGGCCACAGAGGCCAGGCCA		
	ORF Start: ATG at 10	ORF Stop: TGA at 766	
	SEQ ID NO: 10	252 aa	MW at 28126.7kD
NOV2a, CG58542-01 Protein Sequence	MQLTRCCFVFLVQGSLLYVICQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSR PMANSTLLGLLAPPGEAWGILGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNL LVTGKIVDHGNGTFSVHFQHNATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFN CRMEWEKVERGRRTSLCTHDPAKICSRDHAQSSATWSCSQPFKVVVCYVIAFYSTDYRL VQKVC PDYNYHSDTPYYP SG		
	SEQ ID NO: 11	702 bp	
NOV2b, 169679583 DNA Sequence	GGATCCCAGGATGATGGTCTCTCCCGGCTCAGAGGACCCTGAGCGTGATGACCACGAGG GCCAGCCCCGGCCCCGGGTGCTCGGAAGCGGGGCCACATCTACCTAAGTCCCGCCC CATGGCCAATTCACCTCTCTAGGGCTGCTGGCCCCGCCTGGGGAGGCTTGGGGCATT CTTGGGCAGCCCCCAACCGCCGAACCAAGCCCCCACCCTCAGCCAAGGTGAAGA AAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCCCTGAACCTGCT CGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCCACTTCCAACAC AATGCCACAGGCCAGGGAACATCTCCATCAGCCTCGTGCCCCCAGTAAAGCTGTAG AGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAATCTTCAACTG CCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGGCGACCTCGCTCTGCACCCACGAC CCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTGGAGCTGCTCCC AGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCAGGACTATCGGCTGGT CCAGAAGGTGTGCCAGATTACAACATACCATAGTGATACCCCTACTACCATCTGGG CTCGAG		
	ORF Start: GGA at 1	ORF Stop: 5 at 703	
	SEQ ID NO: 12	234 aa	MW at 26037.0kD
NOV2b, 169679583 Protein Sequence	GSQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGI LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFSVHFQ NATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMEWEKVERGRRTSLCTHD PAKICSRDHAQSSATWSCSQPFKVVVCYVIAFYSTDYRLVQKVC PDYNYHSDTPYYP SG LE		
	SEQ ID NO: 13	702 bp	
NOV2c, 169679634 DNA Sequence	GGATCCCAGGATGATGGTCTCTCCCGGCTCAGAGGACCCTGAGCGTGATGACCACGAGG GCCAGCCCCGGCCCCGGGTGCTCGGAAGCGGGGCCACATCTACCTAAGTCCCGCCC CATGGCCAATTCACCTCTCTAGGGCTGCTGGCCCCGCCTGGGGAGGCTTGGGGCATT CTTGGGCAGCCCCCAACCGCCGAACCAAGCCCCCACCCTCAGCCAAGGTGAAGA AAATCTTTGGCTGGGGCGACTTCTACTCCAACATCAAGACGGTGGCCCTGAACCTGCT CGTCACAGGGAAGATTGTGGACCATGGCAATGGGACCTTCAGCGTCCACTTCCAACAC AATGCCACAGGCCAGGGAACATCTCCATCAGCCTCGTGCCCCCAGTAAAGCTGTAG AGTTCCACCAGGAACAGCAGATCTTCATCGAAGCCAAGGCCTCCAAATCTTCAACTG CCGGATGGAGTGGGAGAAGGTAGAACGGGGCCGGCGACCTCGCTTTCACCCACGAC		

	CCAGCCAAGATCTGCTCCCGAGACCACGCTCAGAGCTCAGCCACCTGGAGCTGCTCCC AGCCCTTCAAAGTCGTCTGTGTCTACATCGCCTTCTACAGCAGGACTATCGGCTGGT CCAGAAGGTGTGCCAGATTACAACTACCATAGTGATACCCCTACTACCCATCTGGG CTCGAG		
	ORF Start: GGA at 1	ORF Stop: 5 at 703	
	SEQ ID NO: 14	234 aa	MW at 26037.0kD
NOV2c, 169679634 Protein Sequence	GSQDDGPPGSEDPERDDHEGQPRPRVPRKRGHISPKSRPMANSTLLGLLAPPGEAWGI LGQPPNRPNHSPPPSAKVKKIFGWGDFYSNIKTVALNLLVTGKIVDHGNGTFSVHFQH NATGQGNISISLVPPSKAVEFHQEQQIFIEAKASKIFNCRMWEKVERGRRTSLCTHD PAKICSRDHAQSSATWSCSQPFKVVVCVYIAFYSTDYRLVQKVC PDYNYHSDTPYYP SGLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b through NOV2c.		
Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	23..252	218/230 (94%)
	3..232	218/230 (94%)
NOV2c	23..252	218/230 (94%)
	3..232	218/230 (94%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a	
PSort analysis:	0.7666 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 23 and 24

- 5 A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU29174	Human PRO polypeptide sequence #151 - Homo sapiens, 252 aa. [WO200168848-A2, 20-SEP-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154

AAM39340	Human polypeptide SEQ ID NO 2485 - Homo sapiens, 252 aa. [WO200153312-A1, 26-JUL-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAB87571	Human PRO1327 - Homo sapiens, 252 aa. [WO200116318-A2, 08- MAR-2001]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAB66150	Protein of the invention #62 - Unidentified, 252 aa. [WO200078961-A1, 28-DEC-2000]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154
AAY99401	Human PRO1327 (UNQ687) amino acid sequence SEQ ID NO:218 - Homo sapiens, 252 aa. [WO200012708-A2, 09-MAR-2000]	1..252 1..252	252/252 (100%) 252/252 (100%)	e-154

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q91VX5	SIMILAR TO NEUREXOPHILIN 3 - Mus musculus (Mouse), 252 aa.	1..252 1..252	243/252 (96%) 246/252 (97%)	e-148
Q9Z2N5	Neurexophilin 3 precursor - Rattus norvegicus (Rat), 252 aa.	1..252 1..252	242/252 (96%) 246/252 (97%)	e-148
O95157	Neurexophilin 3 - Homo sapiens (Human), 221 aa (fragment).	32..252 1..221	221/221 (100%) 221/221 (100%)	e-134
P58417	Neurexophilin 1 precursor - Homo sapiens (Human), 271 aa.	79..252 97..271	114/175 (65%) 143/175 (81%)	7e-68
Q63366	Neurexophilin 1 precursor (Neurophilin) - Rattus norvegicus (Rat), 271 aa.	79..252 97..271	114/175 (65%) 143/175 (81%)	7e-68

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

**EXAMPLE 3.**

The NOV3 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 15	1173 bp	
NOV3a, CG58540-01 DNA Sequence	GCCTGCATCATGGAACCTTTTCTAATGCAAGTGGTACTTTTGCCATACGCCTTTTA AAGATACTGTGTCAAGATAACCCCTTCGCACAACGTGTTCTGTTCTCCTGTGAGCATCT CCTCTGCCCTGGCCATGGTTCTCCTAGGGGCAAAGGGAAACACCGCAACCCAGATGGC CCAGATAGAGTCTCTGCTCTGTCAACCCAGGCTGGAGTGCAGACATTATCGGGCTTTC CAGTCGCTTCTCACTGAAGTGAACAAGGCTGGCACACAGTACCTGCTGAGAACGGCCA ACAGGCTCTTTGGAGAGAAAACCTGTCAAGTTCCTCTCAACGTTTAAGGAATCCTGTCT TCAATTCTACCATGCTGAGCTGAAGGAGCTTTCCTTTATCAGAGCTGCAGAAGAGTCC AGGAAACACATCAACACCTGGGTCTCAAAAAAGACCGAAGGTAATAAGAGTTGTG TGCCGGGTAGTCAATTGATGCAGAAACAGGCTGGTCTTGTGAATGCTGTCTATT CAGAGGAAACTGGGATGAACAGTTTGACAAGGAGAACACCGAGGAGAGACTGTTTAAA GTCAGCAAGGCGAGTAAGGAGGAGAAACCTGTGCAATGATGTTTAAGCAATCTACTT TTAAGAAGACCTATATAGGAGAAATATTTACCCAAATCTTGGTGCTTCCATATGTTGG CAAGGAACTGAATATGATCATCATGCTTCCGACGAGACCACTGACTTGAGAACGGTG GAAAAAGTCTCACTTTTGAAGAACTCACAGCCTGGACCAAGCCAGACTGTATGAAGA GTACTGAGGTGAAGTTCTCCTTCCAAAATTTAAACTACAAGAGGATTATGACATGGA ATCTGTGCTTCGGCATTGGGAATTGTTGATGCCTTCCAACAGGGCAAGGCTGACTTG TCGGCAATGTGAGCGAGAGAGACCTGTGTCTGTCCAAGTTCGTGCACAAGAGTTTGG TGGAGGTGAATGAAGAAGGCACCGAGGCAGCGGCAGCGTCGAGCTGCTTTGTAGTTGC AGAGTGCTGCATGGAATCTGGCCCCAGGTTCTGTGCTGACCACCTTTCTTTTCTTC ATCAGGCACAACAGAGCCAACAGCATTCTGTTCTGTGGCAGGTTCTCATCGCCATAAA GGTGCACTTACC		
	ORF Start: ATG at 11	ORF Stop: TAA at 1157	
	SEQ ID NO: 16	382 aa	MW at 43163.1kD
NOV3a, CG58540-01 Protein Sequence	METLSNASGTFAIRLLKILQDNP SHNVFCSPVSISSALAMVLLGAKGNTATQMAQIE SLLCHPGWSADIHRAFQSLLEVNKAGTQYLLRTANRLFGEKTCQFLSTFKESCLQFY HAELKELSFIRAAEESRKHINTWVSKKTEGKIEELLPGSSIDAETRLVLVNAVYFRGN WDEQFDKENTEERLPKVSASKEEKPQMMFKQSTFKKTYIGEIFTQILVLPYVGKEL NMIIMLPDETTDLRTVEKSLTFEKLTAWTKPDCMKSTEVEVLLPKFKLQEDYDMESVL RHLGIVDAFQQGKADLSAMSAERDLCLSKFVHKSFEVNEEGTEAAAASSCFVVAECC MESGPRFCADHPFLFFIRHNRANSILFCGRFSSP		

Further analysis of the NOV3a protein yielded the following properties shown in Table

5 3B.

Table 3B. Protein Sequence Properties NOV3a	
PSort analysis:	0.6881 probability located in mitochondrial inner membrane; 0.6500 probability located in plasma membrane; 0.3773 probability located in mitochondrial intermembrane space; 0.3157 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 3C.

Table 3C. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAV55841	Human cytoplasmic antiproteinase-3 protein (CAP-3) - Homo sapiens, 376 aa. [WO9957273-A2, 11-NOV-1999]	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
AAR99254	Cytoplasmic antiproteinase-3 protein - Homo sapiens, 376 aa. [WO9624650-A2, 15-AUG-1996]	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
AAU30834	Novel human secreted protein #1325 - Homo sapiens, 566 aa. [WO200179449-A2, 25-OCT-2001]	1..382 191..566	324/382 (84%) 351/382 (91%)	0.0
AAB11125	Human thrombin inhibitor protein - Homo sapiens, 376 aa. [US6133422-A, 17-OCT-2000]	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153
AAB59176	Thrombin inhibitor protein - Unidentified, 376 aa. [US6156540-A, 05-DEC-2000]	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3D.

Table 3D. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P50453	Cytoplasmic antiproteinase 3 (CAP3) (CAP-3) (Protease inhibitor 9) (Serpin B9) - Homo sapiens (Human), 376 aa.	1..382 1..376	328/382 (85%) 353/382 (91%)	0.0
Q96J44	SERINE (OR CYSTEINE) PROTEINASE INHIBITOR, CLADE B (OVALBUMIN), MEMBER 6 - Homo sapiens (Human), 376 aa.	1..382 1..376	279/382 (73%) 314/382 (82%)	e-153
P35237	Placental thrombin inhibitor (Cytoplasmic antiproteinase) (CAP) (Protease inhibitor 6) - Homo sapiens (Human), 376 aa.	1..382 1..376	278/382 (72%) 312/382 (80%)	e-152
O02739	Serine proteinase inhibitor B-43 - Bos taurus (Bovine), 378 aa.	1..382 1..378	252/382 (65%) 303/382 (78%)	e-139
Q60854				e-136

	(SERINE (OR CYSTEINE) PROTEINASE INHIBITOR, CLADE B (OVALBUMIN), MEMBER 6) - Mus musculus (Mouse), 378 aa.	1..378	301/383 (78%)	
--	---	--------	---------------	--

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3E.

Table 3E. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin: domain 1 of 1	1..382	170/400 (42%) 314/400 (78%)	8.8e-159

#### EXAMPLE 4.

The NOV4 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 17	502 bp	
NOV4a, CG56340-03 DNA Sequence	GCAATATTGGCAACATCCCAATGGCCCTGTCCTTTTCTTTACTGATGGCCGTGCTGGT GCTCAGCTACAAATCCATCTGTTCTCTGGGCTGTGATCTGCCTCAGACCCACAGCCTG GGTAATAGGAGGGCCTTGATACTCCTGGCACAAATGGGAAGAATCTCTCCTTTCTCCT GCCTGAAGGACAGACATGACTTTGGATTCCCCAGGAGGAGTTTGATGGCAACCAGTT CCAGAAGGCTCAAGCCATCTCTGCTCCTCCATGAGATGATCCAGCAGACCTTCAATCTC TTCAGCACAAAGGACTCATCTGCTACTTGGGAACAGAGCCTCCTAGAAAAATTTTCCA CTGAACTTAACCAGCAGCTGACAGAGAAGAAATACAGCCCTTGTGCTGGGAGGTTGT CAGAGCAGAAATCATGAGATCCTTCTCTTTATCAAAAATTTTCAAGAAAGATTAAGG AGGAAGGAATGAAACCTGTTTCAACATGGAATGATCT		
	ORF Start: ATG at 21	ORF Stop: TGA at 474	
	SEQ ID NO: 18	151 aa	MW at 17402.8kD
NOV4a, CG56340-03 Protein Sequence	MALSFSLMAVLVLSYKISCSLGCPLPQTHSLGNRRALILLAQMGRISPFSLKDRHD FGFPQEEFDGNQFQKAQISVLHEMIQQTFLNFSTKSSATWEQSLLEKFSTELNQL TEKKYSPCAWEVVRAEIMRSFSLSKIFQERLRRKE		
	SEQ ID NO: 19	396 bp	
NOV4b, 174308150 DNA Sequence	GGATCCTGTGATCTGCCTCAGACCCACAGCCTGGGTAATAGGAGGGCCTTGATACTC TGGCACAAATGGGAAGAATCTCTCCTTTCTCCTGCCTGAAGGACAGACATGACTTTGG ATTCCCCAGGAGGAGTTTGATGGCAACAGTCCAGAAGGCTCAAGCCATCTCTGTG CTCCATGAGATGATCCAGCAGACCTTCAATCTCTTCAGCACAAAGGACTCATCTGCTA CTTGGGAACAGAGCCTCCTAGAAAAATTTTCACTGAACTTAACCAGCAGCTGACAGA GAAGAAATACAGCCCTTGTGCTGGGAGGTTGTGAGAGCAGAAATCATGAGATCCTTC TCTTTATCAAAAATTTTCAAGAAAGATTAAGGAGGAAGGAACCTCGAG		
	ORF Start: GGA at 1	ORF Stop: at 397	
	SEQ ID NO: 20	132 aa	MW at 15360.3kD
NOV4b, 174308150 Protein	GSCDLPQTHSLGNRRALILLAQMGRISPFSLKDRHDFGFPQEEFDGNQFQKAQISV LHEMIQQTFLNFSTKSSATWEQSLLEKFSTELNQLTEKKYSPCAWEVVRAEIMRSF SLSKIFQERLRRKELE		

Sequence	
----------	--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

Table 4B. Comparison of NOV4a against NOV4b and NOV4c.		
Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	24..151	128/128 (100%)
	3..130	128/128 (100%)

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

Table 4C. Protein Sequence Properties NOV4a	
PSort analysis:	0.5231 probability located in outside; 0.1317 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

- 5 A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 4D.

Table 4D. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAP20108	Sequence encoded by leukocyte interferon LeIF F cDNA - Homo sapiens, 189 aa. [GB2079291-A, 20-JAN-1982]	1..151 1..189	151/189 (79%) 151/189 (79%)	7e-77
AAP40123	Sequence encoded by the cDNA insert of the recombinant plasmid CG-pBR 322/HLycIFN-1'b - Homo sapiens, 189 aa. [EP100561-A, 15-FEB-1984]	1..151 1..189	150/189 (79%) 150/189 (79%)	3e-76
AAP30179	Sequence of a polypeptide with human lymphoblastoid interferon activity encoded by plasmid CG-pBR 322/HL gamma cIFN-1'b - Homo sapiens, 189 aa. [EP76489-A, 13-APR-1983]	1..151 1..189	150/189 (79%) 150/189 (79%)	3e-76



AAB49780	Human interferon alpha-f amnio acid sequence - Homo sapiens, 189 aa. [WO200107608-A1, 01-FEB-2001]	1..151 1..189	141/189 (74%) 145/189 (76%)	2e-71
AAR62368	Interferon alpha consensus sequence - Synthetic, 187 aa. [WO9420122-A, 15-SEP-1994]	1..151 1..187	139/189 (73%) 144/189 (75%)	7e-69

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
E968396	ARTIFICIAL SEQUENCE FOR CDNA INSERT OF RECOMBINANT PLASMID CG-PBR 322/HLYCIFN-1'B - vectors, 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
E968985	POLYPEPTIDE FOR THE USE OF IMMUNOMODULATOR, ANTI-TUMOR-AGENT - vectors, 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
P01568	Interferon alpha-21 precursor (Interferon alpha-F) (LeIF F) - Homo sapiens (Human), 189 aa.	1..151 1..189	151/189 (79%) 151/189 (79%)	3e-76
CAA00629	ARTIFICIAL SEQUENCE FOR CDNA INSERT OF RECOMBINANT PLASMID CG-PBR 322/HLYCIFN-1'B - synthetic construct, 189 aa.	1..151 1..189	150/189 (79%) 150/189 (79%)	1e-75
Q14608	LEUKOCYTE INTERFERON-ALPHA - Homo sapiens (Human), 181 aa.	9..151 1..181	143/181 (79%) 143/181 (79%)	3e-72

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a			
Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
interferon: domain 1 of 2	1..115	81/116 (70%) 109/116 (94%)	4.9e-71

interferon: domain 2 of 2	116..151	27/36 (75%) 33/36 (92%)	1e-19
---------------------------	----------	----------------------------	-------

**EXAMPLE 5.**

The NOV5 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 21	203 bp	
NOV5a, CG58514-01 DNA Sequence	ACCTCTTTGCCACCATACCATGAAGGTATGCGTGATTGTCTGTCTCTCCTCGTGATA ATAGCCGCCTTCTGCTCTGTAGCACTCTCAGCACCGAATTCCAAACCAAAAGAGGCCAA GCAAGTCTGCGCTGACCCAGTGAGTCTGGGTCCAGGAGTACGTGTATGACCTGGAA CTGAAGTCTGAGCTGCTCAGAGACAGGAAGT		
	ORF Start: ATG at 20	ORF Stop: TGA at 176	
	SEQ ID NO: 22	52 aa	MW at 5408.4kD
NOV5a, CG58514-01 Protein Sequence	MKVCVIVLSLLVIIAFCVSVALSAPNSKPKEASKSALTPVSPGSRSTCMTWN		

Further analysis of the NOV5a protein yielded the following properties shown in Table

5 5B.

Table 5B. Protein Sequence Properties NOV5a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 5C.

Table 5C. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

10 In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5D.

Table 5D. Public BLASTP Results for NOV5a
---

Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
B60407	monocyte adherence-induced protein 5 alpha - human, 52 aa.	1..52 1..52	43/52 (82%) 48/52 (91%)	4e-19

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### EXAMPLE 6.

The NOV6 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis		
	SEQ ID NO: 23	2305 bp
NOV6a, CG57887-01 DNA Sequence	ATTTTTCCTCGGCTGCCGGCGGCTCCGACATCATGCTCCGGCTCCTCCGGCCGCT GCTGCTACTGCTGCTGCTGCCCTCCCCGGGGTCCCCTGAGCCCCCGGCCTGACCCAG CTGTCCCCGGGGCGCCCCCGCAGGCCCCGACTTGCTCTACGCTGACGGGCTGCGCG CCTACGCGGCCGGGGCTTGGGCGCCCGGCCGCTGGCGCTGCTGCGGGAGGCGCTGCGGAG CCAGGCGGCGCTGGGCGGGGTGCGGCTGGATTGCGGGGCGAGCTGCGCGGCCGATCCG GCGCGCGCTCCCCGCGTGCTTCTCGGGGCCCGGAGCCCGACTCCGGGCCGGGAC CCACGCAGGGGTCTGGGAGCGACAGCTTCTCCGTGCAGCGCTCCGCCGCGCAGACTG CCTGACCCAGTGCGCAGCAGGAGGCTGGGCCCCGGGGCGCGGCGGGCTTCCGCGTG GGGAGCGCGCTCCGGGACGCCTTCCGCCGTCCGGAGCCCTACAACCTACCTGCAGAGGG CCTATTACAGTTGAAGAAGCTGGATCTGGCAGCTGCGGCAGCACACACCTTCTTTGT AGCAAACCCATGCACCTGCAGATGCGGGAGGACATGGCTAAGTACAGACGAATGTGCG GGAGTTCCGGCCCAGAGCTTCCGGGACCTGGAGACGCCCCACACTGGGCAGCCTATG AACTGGCCTGGAGCTACTGGGGCGCAGGAGGCGAGGCTGGCACTGCCAGGCTAGA GGAGGCTCTTCAGGGGAGCCTGGCCAGATGGAGAGCTGCCGTGCTGACTGTGAGGGG CCTGAGGAGCAGCAGGGGGCTGAAGAAGAGGAGGATGGGGCTGCGAGCCAGGGGGGCC TCTATGAGGCCATTGCAGGACACTGGATTCAAGTCTGCACTGCCGCAACGCTGTGT GGGGGAAGCAGCCACACGCCCTGGTGCAGCTTCCCTGTCCAGACTTCTTCCCAAC CAGCTGAGGCGCTACATGAGGCCCATGCTCAGGTGGGCAATCTGTCCAGGCTATAG AAAATGTCTGAGTGCTCTGCTCTTCTACCCGAGGATGAGGCTGCCAAGAGGGCTCT GAACCAATACAGGCCAGCTGGGAGAGCCGAGACCTGGCCTCGGACCCAGAGAGGAC ATCCAGCGCTTCTATCCTCCGATCCCTGGGGGAGAAGAGGCAGCTCTACTATGCCATGG AGCACCTGGGGACAGCTTCAAGGATCCTGACCCCTGGACCCCTGCAGCTCTCATCCC TGAGGCACTTAGAGAAAAGCTCAGAGAGGATCAAGAGAAAGAGGCCTTGGGACCATGAG CCCGTGAAGCCAAAGCCCTTGACCTACTGGAAGGATGTCTTCTCTGGAGGGTGTGA CCTTGACCCAGGATTCCAGGCAGCTGAATGGGTCCGAGCGGGCGGTGTGGATGGGCT GCTCACCCCAGCCGAGTGTGGGGTGTGCTGCAGCTGGCTAAGGATGCAGCTGGGGCT GGAGCCAGGTCTGGCTATCGTGGTCCGCGCTCCCTCACACCCCCATGAACGCTTCG AGGGGCTCACGCTGCTTAAGGCTGCGCAGCTGGCCCGGGTGGGACAGTGGGCACTCA GGGTGCTAAGCTGCTTCTGGAGGTGAGCGAGCGGGTGGGACCTTGACCCAGGCCTAC TTCTCCCCGAACGGCCCTGCATCTGTCTTACCCACCTGGTGTGCCGAGCGCCA	

	TAGAAGGAGAGCAAGAGCAGCGCATGGACCTGAGTCACCCAGTGCACGCAGACAACCTGCGTCCTGGACCCTGACACGGGAGAGTGCTGGCGGGAGCCCCAGCCTACACCTATCGG GACTACAGCGGACTCCTCTACCTCAACGATGACTTCCAGGGTGGGGACCTGTTCTTCA CGGAGCCCAACGCCCTCACTGTACGGCTCGGGTGCCTCTCGTGTGGGCGCCTTGT GGCCTTCAGCTCCGGTGTGAGAATCCCCATGGGGTGTGGGCCGTGACTCGGGGACGG CGCTGTGCCCTGGCACTGTGGCACACGTGGGCACCTGAGCACAGGGAGCAGGAGTGGG TAGAAGCCAAAGAACTGCTGCAGGAGTACAGGAGGAGGAGGAAGAGGAAGAGGAAGA AATGCCAGCAAAGACCCTTCCCAGAGCCCCCTAGCCGAGGACACAGAGGGTCCAA GACAAGACTGGAAGGGCACCTCGGGTTCGGGAGGAGCTGTGAGTGCTGAGCCAGCTC CTTGAGGATGTGGCCACTTGACTTGTGGAAGCCATCTTGATG		
	ORF Start: ATG at 36	ORF Stop: TGA at 2244	
	SEQ ID NO: 24	736 aa	MW at 81805.5kD
NOV6a, CG57887-01 Protein Sequence	MLRLLRPLLLLLLLPPGPSPEPPGLTQLSPGAPPQAPDLLYADGLRAYAAGAWAPAVA LLREALRSQAALGRVRLDCGASCAADPGAALPAVLLGAPEPDSGPGPTQGSWERQLLR AALRRADCLTQCAARRLGPGGAAARLRVGSALRDAFRRREPYNLQRAYYQLKLDLAA AAAHTFFVANPMHLQMRDMAKYRRMSGVRPQSFRLDLETPPHWAA YDTGLELLGRQEA GLALPRLEEALQGS LAQMESC RADCEGP EEQQGA EEEEDGAASQGGLYEAIAGHWIQV LQCRQRCVGEAATRPGRSFPVPDFLPNQLRRLHEAHAQVGNLSQAIENVLSVLLFYPE DEAAKRALNQYQAQLGEPRPGLGPREDIQRFILRSLGEKRQLYYAMEHLGTSFKDPDP WTPAALIPEALREKLREDQEKRPWDHEPVKPKPLTYWKDVLLLEGVLTLTQDSRQLNGS ERAVLDGLLTPAECGVLLQLAKDAAGAGARSGYRGRSPHTPHERFEGLTVLKAQLA RAGTVGSQGA LLLLEVSRVRTLTQAYFSPERPLHLSFTHLVCRSAIEGEQEQRMDLS HPVHADNCVLD PDTGECWREPPAYTYRDYSGLLYLNDDFQGGDLFFTEPNALTVTARV RPRCGR LVAFSSGVENPHGVWAVTRGRRCALALWHTWAPEHREQEWEIAKELLQESQE EEEEEEEEMPSKDPSPPEPPSRRHQRVQDKTGRAPRVREEL		

Further analysis of the NOV6a protein yielded the following properties shown in Table 6B.

Table 6B. Protein Sequence Properties NOV6a	
PSort analysis:	0.4991 probability located in lysosome (lumen); 0.3700 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 21 and 22

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 6C.

Table 6C. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93142	Human protein sequence SEQ ID NO:12045 - Homo sapiens, 736 aa. [EP1074617-A2, 07-FEB-2001]	37..714 35..720	314/706 (44%) 421/706 (59%)	e-162
AAB93215				e-161

	NO:12194 - Homo sapiens, 736 aa. [EP1074617-A2, 07-FEB-2001]	35..720	421/706 (59%)	
AAB88373	Human membrane or secretory protein clone PSEC0109 - Homo sapiens, 736 aa. [EP1067182-A2, 10-JAN-2001]	37..714 35..720	313/706 (44%) 421/706 (59%)	e-161
AAB36392	Human tumour suppressor Gros1-S protein SEQ ID NO:4 - Homo sapiens, 736 aa. [WO200065047-A1, 02-NOV- 2000]	37..714 35..720	312/706 (44%) 419/706 (59%)	e-160
AAB36393	Mouse tumour suppressor Gros1-L protein SEQ ID NO:6 - Mus musculus, 747 aa. [WO200065047- A1, 02-NOV-2000]	24..714 22..722	308/721 (42%) 424/721 (58%)	e-159

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6D.

Table 6D. Public BLASTP Results for NOV6a				
Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13512	PROTEIN B - Homo sapiens (Human), 551 aa.	186..736 1..551	551/551 (100%) 551/551 (100%)	0.0
Q15740	CHROMOSOME 12P13 SEQUENCE, COMPLETE SEQUENCE (HYPOTHETICAL 62.3 KDA PROTEIN) - Homo sapiens (Human), 551 aa.	186..736 1..551	550/551 (99%) 550/551 (99%)	0.0
O88836	CHROMOSOME 6 BAC-284H12 (RESEARCH GENETICS MOUSE BAC LIBRARY) COMPLETE SEQUENCE (RESEARCH GENETICS MOUSE BAC LIBRARY) (GENE RICH CLUSTER, B GENE) - Mus musculus (Mouse), 545 aa.	190..736 1..545	477/549 (86%) 508/549 (91%)	0.0
Q96SL5	CDNA FLJ14774 FIS, CLONE NT2RP4000051, WEAKLY SIMILAR TO SYNAPTONEMAL COMPLEX PROTEIN SC65 - Homo sapiens (Human), 736 aa.	37..714 35..720	314/706 (44%) 421/706 (59%)	e-161
Q96SK8	CDNA FLJ14791 FIS, CLONE NT2RP4001064, WEAKLY SIMILAR	37..714 35..720	313/706 (44%) 421/706 (59%)	e-161

	PROTEIN SC65 - Homo sapiens (Human), 736 aa.			
--	---	--	--	--

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6E.

Table 6E. Domain Analysis of NOV6a			
Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### EXAMPLE 7.

The NOV7 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 25	372 bp	
NOV7a, CG57885-01 DNA Sequence	CCATGAACAGCGGCGTGTGCCTGTGTGTGCTGATGGCGGTA CTGGCGGCTGGCGCCCTGACGCAGCCGGTGCCTCCCGCAGATCCCGCGGGCTCCGGGCTGCAGCGGGCAGAGGAGCGCCCCGTAGGCAGCTGAGGGTATCGCAGAGAACGGATGGCGAGTCCCGAGCGCACCTGGGCGCCCTGCTGGCAAGATACATCCAGCAGGCCCGGAAAGGTAAGAATGCTGCCTCCCATCCCTCACTTCTGCCCTTGTTCCCAGGCTCCCGATGCTGACCCTCTTCTCTAGCGCTAGCCTGATGGGGATGACCTCTCTCGGTAGGAAACAAGCAACATGATTCTGGCGGTCCTTTGTAGCAATCTGAGAAGGG		
	ORF Start: ATG at 3	ORF Stop: TGA at 336	
	SEQ ID NO: 26	111 aa	MW at 11598.4kD
NOV7a, CG57885-01 Protein Sequence	MNSGVCLCVLMAVLAAGALTQVPFPADPAGSGLQRAEEAPRRQLRVSRQTDGESRAHIGALLARYIQARKGKNAASPSLTALVPRLPMLTLFSSASLMGMTSLGRKQAT		

Further analysis of the NOV7a protein yielded the following properties shown in Table 7B.

Table 7B. Protein Sequence Properties NOV7a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in microbody (peroxisome)
SignalP analysis:	Likely cleavage site between residues 21 and 22

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 7C.

Table 7C. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE10339	Human cholecystokinin (CCK) - Homo sapiens, 136 aa. [WO200168828-A2, 20-SEP-2001]	1..110 22..129	82/113 (72%) 86/113 (75%)	1e-35
AAB24381	Human procholecystokinin amino acid sequence SEQ ID NO:1 - Homo sapiens, 115 aa. [WO200061192-A2, 19-OCT-2000]	1..110 1..108	82/113 (72%) 86/113 (75%)	1e-35
AAY04729	Rat brain cholecystokinin precursor amidation region - Rattus sp, 105 aa. [WO9910361-A1, 04-MAR-1999]	5..110 1..104	56/106 (52%) 62/106 (57%)	1e-19
AAB24382	Human CCK A amino acid sequence CCK-58 SEQ ID NO:2 - Homo sapiens, 58 aa. [WO200061192-A2, 19-OCT-2000]	46..91 1..41	31/46 (67%) 34/46 (73%)	1e-07

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7D.

Table 7D. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P06307	Procholecystokinin precursor (CCK) - Homo sapiens (Human), 115 aa.	1..110 1..108	82/113 (72%) 86/113 (75%)	4e-35
P23362	Procholecystokinin precursor (CCK) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 115 aa.	1..110 1..108	77/113 (68%) 83/113 (73%)	3e-32
P01356	Procholecystokinin precursor (CCK) - Sus scrofa (Pig), 114 aa.	1..110 1..107	66/113 (58%) 73/113 (64%)	2e-24
Q9DCL5	ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:0610025015, FULL INSERT SEQUENCE - Mus musculus (Mouse), 115 aa.	1..110 1..108	63/113 (55%) 71/113 (62%)	1e-22
P09240	Procholecystokinin precursor (CCK) - Mus musculus (Mouse), 115 aa.	1..110 1..108	62/113 (54%) 69/113 (60%)	2e-21

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7E.

Table 7E. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Gastrin: domain 1 of 1	2..71	37/80 (46%) 64/80 (80%)	7.5e-22

**EXAMPLE 8.**

The NOV8 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 27	479 bp	
NOV8a, CG57865-01 DNA Sequence	TGACTGTATCGCCGGAATTCATGAAGGATCGATTCAAGTGGTTGTCGCTGGAGCTGCT CCTGCTGATAGGCGCCGAGTCGCCTTCCGGACGGCGCTCCGGCGGACACGTGCGGTG AAGCAGCGGGCGAATCAGCCGAATCATGGCAAGGCCCGAGTCAGCCGGCTCACTCGA ATCCGTACGAGGTGGTGGCCGTTGCGCAGACCTACCATCCCGGCCAGCAGATATCGGT GGTCATCTATCCGCACTCGGACCAGAGCACTGTCTTCCGGGGATTCTTCTGCAAGGCG CGCGATGCCAACTCGAACGAGTGGATCGGCGAGTGGGTGCAGAGCGAGAACACCAAGA CCATTCCAGAGTGCTCGGCCATCAGCACTCGGACAACCGGGACAAGCTGGGGGCCAA GCTCATCTGGAAGGCACCGCAAATAAGCGGGGACAAGTCTACTTCACGTAAGTGCAG CCAAGCTAATCCGG		
	ORF Start: ATG at 21	ORF Stop: TAA at 456	
	SEQ ID NO: 28	145 aa	MW at 16248.2kD
NOV8a, CG57865-01 Protein Sequence	MKDRFKWLSLELLLLIGA AVAFPDGAPADTCVKQRANQPNHGKARSQPAHSNPYEVVA VAQTYHPGQQISVVIYPHSDQSTVFRGFFLQARDANSNEWIGEWVQSENTKTIPECSA ITHSDNRDKLGAKLIWKAPQNKRGQVYFT		
	SEQ ID NO: 29	384 bp	
NOV8b, 171651532 DNA Sequence	GGATCCTTTCGGACGGCGCTCCGGCGGACACGTGCGTGAAGCAGCGGGCGAATCAGC CGAATCATGGCAAGGCCCGGAGTCAGCCGGCTCACTCGAATCCGTACGAGGTGGTGGC CGTTGCGCAGACCTACCATCCCGGCCAGCAGATATCGGTGGTCATCTATCCGCACTCG GACCAGAGCACTGTCTTCCGGGGATTCTTCTGCAAGGCGCGCATGCCAACTCGAAGC AGTGGATCGGCGAGTGGGTGTAGAGCGAGAACACCAAGACCATTCCAGAGTGCTCGGC CATCACGCACTCGGACAACCGGGACAAGCTGGGGGCCAAGCTCATCTGGAAGGCACCG CAAATAAGCGGGGACAAGTCTACTTCACGCTCGAG		
	ORF Start: GGA at 1	ORF Stop: TAG at 253	
	SEQ ID NO: 30	84 aa	MW at 9265.1kD
NOV8b, 171651532 Protein Sequence	GSFPDGAPADTCVKQRANQPNHGKARSQPAHSNPYEVVAQTYHPGQQISVVIYPHS DQSTVFRGFFLQARDANSNEWIGEV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 8B.



<b>Table 8B. Comparison of NOV8a against NOV8b and NOV8c.</b>		
<b>Protein Sequence</b>	<b>NOV8a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV8b	21..103 2..84	82/83 (98%) 83/83 (99%)

Further analysis of the NOV8a protein yielded the following properties shown in Table 8C.

<b>Table 8C. Protein Sequence Properties NOV8a</b>	
<b>PSort analysis:</b>	0.6377 probability located in outside; 0.1821 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
<b>SignalP analysis:</b>	Likely cleavage site between residues 22 and 23

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 8D.

<b>Table 8D. Geneseq Results for NOV8a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV8a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

<b>Table 8E. Public BLASTP Results for NOV8a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV8a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9VAN1	CG14515 PROTEIN - Drosophila melanogaster (Fruit fly), 145 aa.	1..145 1..145	144/145 (99%) 144/145 (99%)	6e-82

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8F.

<b>Table 8F. Domain Analysis of NOV8a</b>			
<b>Pfam Domain</b>	<b>NOV8a Match Region</b>		<b>Expect Value</b>

		Similarities for the Matched Region	
Reeler: domain 1 of 1	30..145	31/150 (21%) 78/150 (52%)	2.8e-05

**EXAMPLE 9.**

The NOV9 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis			
	SEQ ID NO: 31	669 bp	
NOV9a, CG54503-03 DNA Sequence	TCCCGCGGGCCAGCGCACTACGAGATGCTGGGTCGCTGCCGCATGGTGTGCGACCCGC ATGGGCCCCGTGGCCCTGGTCCCAGCGCGCGCTGCTTCCGTGCCCCCTTCCCGCC AGGCGCCAAGGGAGAGGTGGGCCGGCGCGGGAAGCAGGCCTGCCGGGGCCCCCTGGA CCACCAGGTCCAAGAGGGCCCCCAGGAGAACCCGGCAGGCCAGGCCCCCGGGCCCTC CCGGTCCAGGTCCGGGCGGGGTGGCGCCCGCTGCCGGCTACGTGCCTCGCATTGCTTT CTACGCGGGCCTGCCGCGGCCCCACGAGGGTTACGAGGTGCTGCGCTTCGACGACGTG GTGACCAACGTGGGCAACGCCTACGAGGCAGCCAGCGGCAAGTTTACTTGCCCCATGC CAGGCGTCTACTTCTTCGCTTACCACGTGCTCATGCGCGGCGGCGACGGCACCAGCAT GTGGCCGACCTCATGAAGAACGGACAGGTCCGGGCCAGCGCCATTGCTCAGGACGCG GACCAGAACTACGACTACGCCAGCAACAGCGTCATTCTGCACCTGGACGTGGGCGACG AGGTCTTCATCAAGCTGGACGGCGGGAAGTGCACGGCGGCAACACCAACAAGTACAG CACCTTCTCCGGCTTCATCATCTACCCCGAC		
	ORF Start: TCC at 1	ORF Stop: th at 670	
	SEQ ID NO: 32	223 aa	MW at 23296.1kD
NOV9a, CG54503-03 Protein Sequence	SRGPAHYEMLGRCRMVCDPHGPRGPGPDGAPASVPPFPPGAKGEVGRRGKAGLRGPPG PPGPRGPPGEPGRPGPPGPPGPGGVAPAAGYVPRIAFYAGLRPPHEGYEVLRFDDV VTNVGNAYEAAAGKFTCPMPGVYFFAYHVLMRGGDGTSMWADLMKNGQVRASAIQDA DQNYDYASNSVILHLDVGDEVFIKLDGGKVHGGNTNKYSTFSGFIIYPD		

Further analysis of the NOV9a protein yielded the following properties shown in Table

5 9B.

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.8276 probability located in lysosome (lumen); 0.4500 probability located in cytoplasm; 0.4128 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 9C.

Table 9C. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a	Identities/	Expect Value

		<b>Match Residues</b>	<b>the Matched Region</b>	
AAG64212	Murine HSP47 interacting protein, #2 - Mus sp, 255 aa. [JP2001145493- A, 29-MAY-2001]	5..223 20..255	169/236 (71%) 183/236 (76%)	1e-91
AAM40913	Human polypeptide SEQ ID NO 5844 - Homo sapiens, 755 aa. [WO200153312-A1, 26-JUL-2001]	19..222 519..754	90/242 (37%) 121/242 (49%)	2e-32
AAM39127	Human polypeptide SEQ ID NO 2272 - Homo sapiens, 744 aa. [WO200153312-A1, 26-JUL-2001]	19..222 508..743	90/242 (37%) 121/242 (49%)	2e-32
AAM40607	Human polypeptide SEQ ID NO 5538 - Homo sapiens, 255 aa. [WO200153312-A1, 26-JUL-2001]	19..223 43..252	82/218 (37%) 112/218 (50%)	3e-30
AAM38821	Human polypeptide SEQ ID NO 1966 - Homo sapiens, 253 aa. [WO200153312-A1, 26-JUL-2001]	19..223 41..250	82/218 (37%) 112/218 (50%)	3e-30

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9D.

<b>Table 9D. Public BLASTP Results for NOV9a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV9a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
O88992	C1q-related factor precursor - Mus musculus (Mouse), 258 aa.	1..223 15..258	171/244 (70%) 183/244 (74%)	2e-93
O75973	C1q-related factor precursor - Homo sapiens (Human), 258 aa.	1..223 15..258	174/244 (71%) 185/244 (75%)	4e-93
Q9ESN4	Gliacolin precursor - Mus musculus (Mouse), 255 aa.	5..223 20..255	169/236 (71%) 183/236 (76%)	5e-91
Q921S8	PROCOLLAGEN, TYPE VIII, ALPHA 1 - Mus musculus (Mouse), 744 aa.	19..222 509..743	94/241 (39%) 123/241 (51%)	3e-34
Q9D2V4	PROCOLLAGEN, TYPE VIII, ALPHA 1 - Mus musculus (Mouse), 744 aa.	19..222 509..743	94/241 (39%) 123/241 (51%)	3e-34

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

<b>Table 9E. Domain Analysis of NOV9a</b>
---

Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Collagen: domain 1 of 1	35..92	36/60 (60%) 49/60 (82%)	0.0043
Clq: domain 1 of 1	96..220	43/140 (31%) 92/140 (66%)	6.4e-29

### EXAMPLE 10.

The NOV10 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 10A.

[illegible]

YVFS DGYL I N Y D L T F L T M K T R L P R P P T R R P S G A H A P P K P V K P N E A S R P
---

Further analysis of the NOV10a protein yielded the following properties shown in Table 10B.

Table 10B. Protein Sequence Properties NOV10a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 10C.

Table 10C. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y54368	Protein encoded by colon specific gene (CSG) clone 2348122 - Homo sapiens, 510 aa. [WO9960161-A1, 25-NOV-1999]	5..480 32..506	219/482 (45%) 310/482 (63%)	e-114
AA Y22201	Human extracellular mucous matrix glycoprotein protein sequence - Homo sapiens, 510 aa. [US5929033-A, 27-JUL-1999]	5..480 32..506	219/482 (45%) 310/482 (63%)	e-114
AA E03653	Human extracellular matrix and cell adhesion molecule-17 (XMAD-17) - Homo sapiens, 510 aa. [WO200142285-A2, 14-JUN-2001]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114
AA B50955	Human PRO698 protein - Homo sapiens, 510 aa. [WO200073348-A2, 07-DEC-2000]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114
AA B65169	Human PRO698 (UNQ362) protein sequence SEQ ID NO:67 - Homo sapiens, 510 aa. [WO200073454-A1, 07-DEC-2000]	10..480 35..506	217/477 (45%) 308/477 (64%)	e-114

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10D.

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H1L6	BA209J19.1.1 (GW112 PROTEIN) - Homo sapiens (Human), 510 aa.	10..480 35..506	217/477 (45%) 308/477 (64%)	e-113
Q07081	Olfactomedin precursor (Olfactory mucus protein) - Rana catesbeiana (Bull frog), 464 aa.	53..477 32..460	155/441 (35%) 247/441 (55%)	2e-68
AAL66227	NOELIN-1 - Xenopus laevis (African clawed frog), 485 aa.	28..478 48..475	114/458 (24%) 194/458 (41%)	1e-32
AAL66226	NOELIN-2 - Xenopus laevis (African clawed frog), 458 aa.	32..478 25..448	113/454 (24%) 192/454 (41%)	3e-32
O95362	GW112 PROTEIN - Homo sapiens (Human), 187 aa.	139..315 2..176	77/178 (43%) 107/178 (59%)	7e-32

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10E.

Table 10E. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
OLF: domain 1 of 1	224..481	93/294 (32%) 170/294 (58%)	8.1e-72

#### EXAMPLE 11.

The NOV11 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis		
	SEQ ID NO: 35	1134 bp
NOV11a, CG57572-01 DNA Sequence	GCAGAAGAATAGGCTACTTTATTTCTGAAAAGGAGGGAGTTCCTGCCACCCATTGCA GGGAGGTGCGCCATCAGGACAGTGAAGATGGTGACCCTGCGGAAGAGGACCTGAAAGT GCTCACCTTCCTCGTGCTCTTCATCTTCCTCACCTCCTTCCTGAAGTACTCCACGCC ATGGTGGCCACCACTGGTCCCAAGAAGATGGCCCTGGAGCTCTTGGAGAACCTGA AGAGACTGATCAAGCACAGGCCCTGCACTTGCAACCACTGCATCAGGCAGCATGGGCT CTCAGCCTGGTTCGATGAGAGGTTCAACCAGATAGTGCAGCTGCTGCTGACTGCCAG AACGCGCTCTTGGAGGACAACACCTACCAATGGTGGCTGAGGCTCCAGCAGGAGAAGA AGCCCAATATCATCAACAATACCATCAAGGAATTGAGAGCAGTACCTGGGAATGTGGA CCCAATGCTGGAGAAGAGGTGGTGGGCTGCTGGCACTGTGCTGCTGGGCAACTCG GGCAACCTGAGGCAATTGTCATATCACAATTTATGCTCAGGATGAACAAGGCACCA CGGCAGGGTTTGAAGCTGCTGCCGGGAGCAAAACCGCCACCATCTGGTGTACCTGA GAGCTTCCGGGAGCTGGGGACAATGTGAGCATGGTCTGGTGCCCTTAAAGACCATG AACTTGGAGTGGGTGGTGAGCACCACCACGGGTGCCATTTCCACACCTACACCC	

	CGGTCCTCGTGAAGATCAGAGTGAAACAGGATAAGATCCTGATCTACCACCCAGCCTT CATCAAGTATGTCTTCGACAACCTGGCTGCAGAGCCACAGGCGGTACCCACTCACCAGC ATCCTCTCGGTATCTTCTCAATGCATGTCTGCGATAAGGTAGACTTGTATAGCTTCG GAGCAGATAGCAAAGGGAAGTGGCACCCTACTGGGAGAACACCTGTCTGCGGGGTC TTTTCAAGACGGGGGTGCACGATGCAGGCTTTGAGTCTAACGTGACGGCCACCTTG GCTTCATCAATAAAATCCCGATCTTCAAGGGGAGATGACACAGTGAAGGGGTGAGGAT GGATGCCCCATCATGCCTCTGCGTTTCAAGCC		
	ORF Start: ATG at 85	ORF Stop: TGA at 1096	
	SEQ ID NO: 36	337 aa	MW at 38559.2kD
NOV11a, CG57572-01 Protein Sequence	MVTLRKRTLKVLTFVLVFLTSFLNYSHAMVATTWFPKMALELLENLKRLIKHRPC TCTHCIRQHGLSAWFDERFNQIVQLLLTAQNALLEDNITYQWWLRLQEQKPNINNTI KEFRAVPGNVDPMLEKRSVGCWHCAVVGNSGNLRQLSYHNFMLRMNKAPTAGFEAAAG SKTAHHLVYPESFRELGDNVSMVLVPLKTMNLEWVVSSTTTGAISHTYTPVLVKIRVK QDKILIYHPAFIKYVFDNWLQSHRRYPLTSILSVIFSMHVC DKVDLYSFGADSKGNWH HYWENNLSAGSFHKTGVHDAGFESNVTATLASSIKSRSSRGDDTVKG		

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.8200 probability located in outside; 0.5054 probability located in lysosome (lumen); 0.1565 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 31 and 32

- A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 11C.

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR65244	Human ST30 sialyltransferase - Homo sapiens, 340 aa. [WO9504816-A, 16-FEB-1995]	1..331 1..339	274/339 (80%) 292/339 (85%)	e-158
AAR65240	Porcine ST30 sialyltransferase - Sus scrofa, 343 aa. [WO9504816-A, 16-FEB-1995]	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
AAR41670	Porcine sialyltransferase - Sus scrofa, 343 aa. [WO9318157-A, 16-SEP-1993]	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
AAR75198	Rat Gal-beta-1,3GalNAc, alpha-2,3-	12..332 12..350	149/341 (43%) 203/341 (58%)	6e-78

	norvegicus, 350 aa. [JP07236477-A, 12-SEP-1995]			
AAR75200	Rat P-F4M active fragment, SF-314R - Rattus norvegicus, 314 aa. [JP07236477-A, 12-SEP-1995]	50..332 26..314	135/290 (46%) 183/290 (62%)	3e-76

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

Table 11D. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q11201	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (GAL-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Homo sapiens (Human), 340 aa.	1..331 1..339	278/339 (82%) 295/339 (87%)	e-160
Q9UN51	ALPHA-2,3-SIALYLTRANSFERASE - Homo sapiens (Human), 340 aa.	1..331 1..339	276/339 (81%) 294/339 (86%)	e-159
P54751	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Mus musculus (Mouse), 337 aa.	4..331 1..336	230/336 (68%) 273/336 (80%)	e-137
A45073	Gal beta 1,3GalNAc alpha 2,3-sialyltransferase - pig, 343 aa.	5..331 6..342	234/337 (69%) 272/337 (80%)	e-137
Q02745	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2, 3-sialyltransferase (EC 2.4.99.4) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (GAL-NAC6S) (GAL-beta-1,3-GALNAC-alpha-2,3-sialyltransferase) (ST3GALIA) (ST3O) (ST3GALA.1) (SIAT4-A) - Sus scrofa (Pig), 343 aa.	5..331 6..342	234/337 (69%) 272/337 (80%)	e-136

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.



Table 11E. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IF3: domain 1 of 1	193..202	6/10 (60%) 9/10 (90%)	6.3
Glyco_transf_29: domain 1 of 1	60..331	97/324 (30%) 223/324 (69%)	4.7e-73

**EXAMPLE 12.**

The NOV12 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis		
	SEQ ID NO: 37	4295 bp
NOV12a, CG57518-01 DNA Sequence	TCTTCGTCGCGCTCTCTCTCACCTCTCAGGGAAAGGGGGGACATAGGGCGCTCG CGGGGCCCCGGCGAATGCGCCCCCGCGCCTCTCGGGCTGCGCCGCTCGCGGGAT GAAGCACCGGCGTGAAGATGGAGGTGACCTGCCTTCTACTTCTGGCGCTGATCCCCCT TCCACTGCCGGGACAAGGAGTCTACGCTCCAGCCAGGCGCAGATCGTGCATGCGGG CCAGGCATGTGTGGTGAAGAGGACAATATCAGCGAGCGTGTCTACACCATCCGGGAG GGGGACACCCTCATGCTGCAGTGCCTTGTAACAGGGCACCCCTCGACCCAGGTACGGT GGACCAAGACGGCAGGTAGCGCCTCGGACAAGTTCAGGAGACATCGGTGTTCAACGA GACGCTGCGCATCGAGCGTATTGCACGCACGCAGGGCGGCCGCTACTACTGCAAGGCT GAGAACGGCGTGGGGTGCCGGCCATCAAGTCCATCCGCGTGACGTGCAGTACCTGG ATGAGCCAATGTGACGGTGACACGAGCGGTGAGCGATGTGCGAGGCAACTTCTACCA GGAGAAGACGGTGTTCCTGCGCTGTACTGTCAACTCCAACCCGCTGCCCGCTTCATC TGGAAGCGGGTTCCGATACCCTATCCACAGCCAGGACAATGGGGTTGACATCTATG AGCCCCCTACACTCAGGGGGAGACCAAGGTCTGAAGCTGAAGAACCCTGCGGCCCA GGACTATGCCAGCTACACCTGCCAGGTGTCTGTGCGTAACGTGTGCGGCATCCAGAC AAGGCCATCACCTTCGGGTCACCAACACCACGGCACCACAGCCCTGAAGCTGTCTG TGAACGAAACTCTGGTGGTGAACCCCTGGGGAGAATGTGACGGTGCAGTGTCTGCTGAC AGGCGGTGATCCCCCTCCCCAGCTGCAGTGGTCCCATGGGCCTGGCCCACTGCCCTG GGTGCTCTGGCCAGGGTGGCACCCCTCAGCATCCCTTCAGTGAGCCCGGACTCTG GCTACTACAAGTGCACAGCCACCAACAATGTGGCAACCCTGCCAAGAAGACTGTCAA CCTGCTGGTGCGATCCATGAAGAAGCTACATTCAGATCACTCCTGACGTGATCAAA GAGAGTGAGAACATCCAGCTGGGCCAGGACCTGAAGCTATCGTGCCACGTGGATGCAG TGCCCCAGGAGAAGGTGACCTACAGTGGTTCAAGAATGGCAAGCCGGCACGCATGTC CAAGCGGCTGCTGGTGACCCGCAATGATCCTGAGCTGCCCGCAGTCACCAGCAGCCTA GAGCTCATTGACCTGCACTTCAGTGACTATGGCACCTACCTGTGCATGGCTTCTTTCC CAGGGGCACCCGTGCCGACCTCAGCGTCGAGGTCAACATCTCCTCTGAGACAGTGCC GCCCACCATCAGTGTGCCAAGGGTAGGGCCGTGGTGACCGTGCGCGAGGGATCGCCT GCCGAGCTGCAATGCGAGGTGCGGGGCAAGCCGCGGCCGAGTGCTCTGGTCCCGCG TGGACAAGGAGGCTGCACTGCTGCCCTCGGGGCTGCCCTGGAGGAGACTCCGGACGG GAAGCTGCGGCTGGAGCGAGTGAGCCGAGACATGAGCGGACCTACCCTGCCAGACG GCCCCTATAATGGCTTCAACGTGCGCCCCGTGAGGCCAGGTGCAGCTGAACGTGC AGTTCGCCCGGAGGTGGAGCCAGTCCCAGGACGTGCGCCAGGCGCTGGGCCGGCC CGTGCTCCTGCGCTGCTGCTGCTGCGAGGCAGCCCCAGCGCATCGCTCGGCTGTG TGGCGTTTCAAAGGGCAGCTGCTGCCGCGCGCCCTGTTGTTCCCGCCGCGCCGAGG CGCCGGATCAGCGGAGCTGCGCCTCGACGCGTAACCTCGGACAGCAGCGGACGCTA CGAGTGACGCTCTCAACGATGTGGCTCGGCTGCCTGCTTCCAGGTCTCCGCC AAAGCCTACAGCCCGAGTATTACTTCGACACCCCCAACCCACCCGACGCACAAGC TGTCGAAGAACTACTCTACGTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTCTGA	

	CCCTGTGCTCAACTACAGACTCAGCATCCGCCAGTTGAACCAGCACAAATGCGGTGGTC AAGGCCATCCCGGTCCGGCGTGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCG ATCTCCGTGTGCCCCACAGCTATGAGGTCCGCCTCACACCTTATACCACCTTCGGGGC TGGTGACATGGCCTCCGCATCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCA GTCCTGGCTCTGAGGACCTGTCTCTGGTCCCAGCAGGGTATCCTCTGCAGAGCCC CACACCTCAGTTCTGACTTGGTTTCCCGCTTGCTTCTCAGCCATCAACTCTCCGAA CCTTTTCAGACAACACCTGCCACTTTGAGGATGAGAAGATCTGTGGCTATACCCAGGAC CTGACAGACAACCTTGACTGGACGCGGCAGAAATGCCCTCACCAGAACCCAAACGCT CCCCAACACTGGTCCCCCACCACATAAGTGGCACCCTGAGGGCTACTACATGTT CATCGAGACATCGAGGCTCGGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCTCT TACAATGCCAGCGCCAAGTTCTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAAC ACATCGGCTCCCTCAACCTCCTGGTGCAGTCCCGGAACAAAGGGGCTTGGACACGCA CGCTGGTCTCTCAGTGGCAATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATC AGCCCCAGTGGGCCCTTCCAGATTATTTTGGAGGGGTTTCGAGGCCCGGCTACCTGG GGGATATGCCATAGATGACGTCACTGAAGAAGGGGAGTGTCCCGGAAGCAGAC GGATCCCAATAAAGGTGCAAGACGGGAAGGAGCTGCCTGCGATGGCCTGAAATTCCAC CTTTCATCCCCTATGGATGACGGAGAGCTTACAGATGACCCTATTGAATGCAAGCACC TTTGGATCCATAGAGTGGACAGTAAAGGTGCTCAGTACATGTTGGCTGAGCTGAACCTG CATAATGTGGCCCCCAGGTTCTTGGTCTTTATGGACGAAGGGCACAAGGTTGGTGAA AAGGACTCCGGGGGCCAGCCCTTCCAAGTTTACTGATTTCTCCTTTTACCCTCATG CTATCCCTGAGAAGATGTCAATAATGCCACGTTACAGGTGGGAAAACCTGAGGCTTAG AGAGGAGGAGGAATCTGCCTACGGTCACACAGCTGCAAGGCTAGAGCTGGGACCAGG AGCTGGTCTCTTAACCGACCACTGAGCTCAAGAGCTTTTCTCTGAGCACAACATGA CCCAAAGTGTGCGCGAGCCTATCACAGGTCCCCTGCAATGCCAAACATACACGCACAG CAATACACAACACCTGGGGACATGGATGAAGCTGGAAACCATATTCTCAGCAAACCTG ACACAAGAACAGAAAACCAAACACCACATGTTCTCACTACCAACCCAGTCTGCCCCGC CCTCTCTCTCTCACTGAACTTCCCCTCTCCTCAAACCTCTCGAGGCCACGCCTCTAT GTCCTTGGATGATGATGATGACGACGACGACGATGATGATGATGATGATGATGATGAT GACAATGATGATGATGATGGAAGGAAGACCTACAGAATCCCTCAGGCTCTGACCTCA GTGCTTGTGGGTGGGTGAATGACCACATGTGCGAGGGAGACTCCACAGGTCTCTCCGA TGAGAAGCACTCTTATGCCAAAGAGGAGACTCAGGCCAACTGACAGGACCAGGAATT AGCTACCTGGTAAACCCAGCTATCGACTGCACCCGAGCGGTACACACCACTGGAGC AGTTCAGGGAGAAAGCCACCGCATGCTCACCCTGATGTCTCTGGCTCTGTTTCCCTC TTTCTGCTTCCCCTTCCCACCTCTGAGTCTCTGTGTTCTGTGTTCTGTGTTTCCCCT TCTGCCTGTCTCTGCCGCTTCTCTCTCTGGGCTGGTCTCTCCGAGACTCTGTTCCCT TGGCTGGCATGCCCTCCACCTCCCCTGATGCTGGAGCAGTTACGGGAGAAAGCCACCG GCATGCTACCGTATGTCTCTGGCTCTGTTTCTTCTGTTTCCCCTTCCCACCT TGA		
	ORF Start: ATG at 135	ORF Stop: TGA at 4293	
	SEQ ID NO: 38	1386 aa	MW at 153195.2kD
NOV12a, CG57518-01 Protein Sequence	MEVTCLLLLALIPFHCRCQGVYAPAAQIVHAGQACVVKEDNISERVYTIREGDTLML QCLVTGHPRPQVRWTKTAGSASDKFQETSVFNELRIERIARTQGGRYCYCAENGVG PAIKSIRVDVQYLDEPMLTVHQTVSDVRGNFYQEKTVFLRCTVNSNPAPRFIWKRGSD TLSHSQDNGVDIYEPLYTQGETKVLKLNLRPQDYASYTCQVSVRNVCGIPDKAITFR LTNTTAPPALKLSVNETLVNPGENVTVQCLLTGGDPLPQLQWSHGPGPLPLGALAQQ GTLSPSVQARDSGYNCTATNNVGNPAKKTVNLLVRSKMNAFTQITPDVKESENIQ LGQDLKLSCHVDVAPQEKVYQWFKNGKPARMSKRLLVTRNDPELPAVTSSELEIDLH FSDYGYLCMASFPGAPVPDLSVEVNISSETVPPTISVPKGRAVVTVREGSPAELQCE VRGKPRPPVLWSRVDKEAALLPSGLPLEETPDGKRLRLERSRDSMTYRCQTARYNGF NVRPREAQVQLNVQFPPEVEPSSQDVRLALGRPVLLRCSLLRGSPQRIASAVWRFKQ LLPPPPVVPAAAEAPDHAEALRLDAVTRDSSGSYECVSNDVGSAAFLQVSAKAYSPE IYFDTNPTRSHKLSKNYSYVLQWTQREPDVDPVLNRYLSRQLNQHNNAVVKAI PVRVEKQQLLEYILDLRVPHSYEVRLTPYTTFGAGDMASRIHYTERQIRWPPVLALRT LSSGPKQGILCRAPHLSSDLVSPALFSAINSPNLSNDNTCHFEDEKICGYTQDLTDFD WTRQNALTQNPKRSPNTGPTDISGTPEGYMFIEISRPRELGDRARLVSPLYNASAK FYCVSFFYHMYKHIGSLNLLVRSRNKGALDTHAWSLSGNKGNVWQQAHPVISPSPGF QIIIFEGVRGPGYLGDIADIDVTLKKGCPRKQTDPNKGARREGAACDGLKFLSSPMD DGELTDDPIECKHLWIHRVDSKGAQYMLAELNCIHVAPRFLVFMDEGHKVGKDSGGQ PFQVYTDFSFYPHAIPEKMSIMPTLQVGKLRLEEEESAYGHTAAKARAGTRSWSLNR		

	PPELKSFSLWNTMQSVREPITGPLQCQTYTHSNTQHLGTWMKLETIILSKLTQEOKT KHHMFSLTTQSAPPSLFSPELPLSSNSRGHASMSLDDDDDDDDDDDDDDDDDDDDDD GRKTYRIPPGSDLACGWVNDHMSQGDSTGPPDEKHSYAKEETQAKLTGPGISYPGKP SYRLHPSGYTPLEQFREKATGMLTPYVSGSVSSFCFFPPTSESLCSAHANSPSACLCP LLSLGWSLRDSVPLAGMPSTSPDAGAVQGESHRAHRMSLALFPLSASPSP		
	SEQ ID NO: 39	906 bp	
NOV12b, 170108372 DNA Sequence	GGTACCCACCAGCCCTGAAGCTGTCTGTGAACGAAACTCTGGTGGTGAACCTGGGG AGAATGTGACGGTGCAGTGTCTGTGCTGACAGGCGGTGATCCCCCCCCAGCTGCAGTG GTCCCATGGGCCTGGCCCACTGCCCCCTGGGTGCTCTGGCCCAAGGTGGCACCCTCAGC ATCCCTTCAGTGCAGGCCCGGACTCTGGCTACTACAAGTGCACAGCCACCAACAATG TGGGCAACCCTGCCAAGAAGACTGTCAACCTGCTGGTGCATCCATGAAGAACGCTAC ATTCCAGATCACTCCTGACGTGATCAAAGAGAGTGAGAACATCCAGCTGGGCCAGGAC CTGAAGCTATCGTGCCACGTGGATGCAGTGCCCCAGGAGAAGGTGACCTACCAGTGGT TCAAGAATGGCAAGCCGGCAGCATGTCCAAGCGGCTGCTGGTGACCCGCAATGATCC TGAGCTGCCCGCAGTCACCAGCAGCCTAGAGCTCATTGACCTGCACTTCAGTGACTAT GGCACCTACCTGTGCATGGCTTCTTTCCAGGGGCACCCGTGCCCGACCTCAGCGTCG AGGTCAACATCTCCTCTGAGACAGTGCCGCCACCATCAGTGTGCCCAAGGGTAGGGC CGTGGTGACCGTGCGCGAGGGATCGCCTGCCGAGCTGCAATGCGAGGTGCGGGGCAAG CCGCGGCCGCCAGTGCCCTGGTCCCGCGTGGACAAGGAGGCTGCACTGCTGCCCTCGG GGCTGCCCTTGAGGAGACTCCGGACGGGAAGCTACGGCTGGAGCGAGTGAGCCGAGA CATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAACGTGCGCCCC CGTGAGGCCCAGGTGCAGCTGAACGTGCAGGAATTC		
	ORF Start: GGT at 1	ORF Stop:	
	SEQ ID NO: 40	302 aa	MW at 32832.1kD
NOV12b, 170108372 Protein Sequence	GTPPALKLSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGLPLGALAQQGGLS IPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKESENIQLGQD LKLSCHVDVAVPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLHFSYD GTYLCMASFPGAPVPDLSVEVNISSETVPPTISVPKGRAVTVREGSPAELQCEVRGK PRPPVPWSRVDKEAALLPSGLPLEETPDGKLRLEVRSDMSGTYRCQTARYNGFNVRP REAQVQLNVQEF		
	SEQ ID NO: 41	906 bp	
NOV12c, 170108393 DNA Sequence	GGTACCCACCAGCCCTGAAGCTGTCTGTGAACGAAACTCTGGTGGTGAACCTGGGG AGAATGTGACGGTGCAGTGTCTGTGCTGACAGGCGGTGATCCCCCCCCAGCTGCAGTG GTCCCATGGGCCTGGCCCACTGCCCCCTGGGTGCTCTGGCCCAAGGTGGCACCCTCAGC ATCCCTTCAGTGCAGGCCCGGACTCTGGCTACTACAAGTGCACAGCCACCAACAATG TGGGCAACCCTGCCAAGAAGACTGTCAACCTGCTGGTGCATCCATGAAGAACGCTAC ATTCCAGATCACTCCTGACGTGATCAAAGAGAGTGAGAACATCCAGCTGGGCCAGGAC CTGAAGCTATCGTGCCACGTGGATGCAGTGCCCCAGGAGAAGGTGACCTACCAGTGGT TCAAGAATGGCAAGCCGGCAGCATGTCCAAGCGGCTGCTGGTGACCCGCAATGATCC TGAGCTGCCCGCAGTCACCAGCAGCCTAGAGCTCATTGACCTGCACTTCAGTGACTAT GGCACCTACCTGTGCATGGCTTCTTTCCAGGGGCACCCGTGCCCGACCTCAGCGTCG AGGTCAACATCTCCTCTGAGACAGTGCCGCCACCATCAGTGTGCCCAAGGGTAGGGC CGTGGTGACCGTGCGCGAGGGATCGCCTGCCGAGCTGCAATGCGAGGTGCGGGGCAAG CCGCGGCCGCCAGTGCTCTGGTCCCGCGTGGACAAGGAGGCTGCACTGCTGCCCTCGG GGCTGCCCTTGAGGAGACTCCGGACGGGAAGCTGCGGCTGGAGCGAGTGAGCCGAGA CATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAACGTGCGCCCC CGTGAGGCCCAGGTGCAGTGAACGTGCAGGAATTC		
	ORF Start: GGT at 1	ORF Stop:	
	SEQ ID NO: 42	302 aa	MW at 32848.2kD
NOV12c, 170108393 Protein Sequence	GTPPALKLSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGLPLGALAQQGGLS IPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKESENIQLGQD LKLSCHVDVAVPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTSSLELIDLHFSYD GTYLCMASFPGAPVPDLSVEVNISSETVPPTISVPKGRAVTVREGSPAELQCEVRGK PRPPVLWSRVDKEAALLPSGLPLEETPDGKLRLEVRSDMSGTYRCQTARYNGFNVRP REAQVQLNVQEF		

	SEQ ID NO: 43	720 bp
NOV12d, 170343246 DNA Sequence	GGTACCTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCTATACCACTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATACCCAGGACCTGACAGACAACCTTGACTGGACGCGGCAGAA TGCCCTCACCAGAACCCAAACGCTCCCCAACACTGGTCCCCCACCAGACATAAGT GGCACCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACTGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACTGAAG AAGGGGAGTGTCCCCGGGAATTC	
	ORF Start: GGT at 1	ORF Stop: at 721
	SEQ ID NO: 44	240 aa MW at 26966.1kD
NOV12d, 170343246 Protein Sequence	GTLNQHNNAVVKAI PVRRVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IH YTEPINSNPNSDNTCHFEDEKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDIS GTPEGYYMFIETSRPRELGDRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLVRS RNKGALDTHAWSLSGNKGNVWQAHVPI SPSPGFQI IFEGVRGPGYLGDI AIDVTLK KGECPPREF	
	SEQ ID NO: 45	720 bp
NOV12e, 170343692 DNA Sequence	GGTACCTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCTATACCACTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCTTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATACCCAGGACCTGACAGACAACCTTGACTGGACGCGGCAGAA TGCCCTCACCAGAACCCAAACGCTCCCCAACACTGGTCCCCCACCAGACATAAGT GGCACCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCTGCGCCAAGTTCTACTGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACTGAAG AAGGGGAGTGTCCCCGGGAATTC	
	ORF Start: GGT at 1	ORF Stop: at 721
	SEQ ID NO: 46	240 aa MW at 26998.2kD
NOV12e, 170343692 Protein Sequence	GTLNQHNNAVVKAI PVRRVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IH YTEPINSNPNSDNTCHFEDEKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDIS GTPEGYYMFIETSRPRELGDRARLVSPLYNACAKFYCVSFFYHMYGKHIGSLNLLVRS RNKGALDTHAWSLSGNKGNVWQAHVPI SPSPGFQI IFEGVRGPGYLGDI AIDVTLK KGECPPREF	
	SEQ ID NO: 47	720 bp
NOV12f, 170684238 DNA Sequence	GGTACCTGAACCAGCACAAATGCGGTGGTCAAGGCCATCCCGGTCCGGCGTGTGGAGA AGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGCTATGAGGT CCGCCTCACACCTATACCACTTCGGGGCTGGTGACATGGCCTCCCGCATCATCCAC TACACAGAGCCCATCAACTCTCCGAACCTTTCAGACAACACCTGCCACTTTGAGGATG AGAAGATCTGTGGCTATACCCAGGACCTGACAGACAACCTTGACTGGACGCGGCAGAA TGCCCTCACCAGAACCCAAACGCTCCCCAACACTGGTCCCCCACCAGACATAAGT GGCACCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGG ACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACCGTGTCTC CTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCC CGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATG TGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCAGATTATTTTGA GGGGGTTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACTGAAG	

	AAGGGGGAGTGTCCCCGGGAATTC		
	ORF Start: GGT at 1	ORF Stop: at 721	
	SEQ ID NO: 48	240 aa	MW at 27035.1kD
NOV12f, 170684238 Protein Sequence	GTLNQHNNAVVKAI PVR RVEKGQLLEYILTDLRVPHSYEVRLTPYTTFGAGDMASRI IHYTEPINSPLNSDNTCHFEDEKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELGDRARLVSPLYNASAKFYRVVSFFYHMYGKHIGSLNLLVRSRNKGALDTHAWSLSGNKGNVWQAHVPI SPSPGFQII FEGVRGPGYLGDI AIDDVTLKKGECPRF		
	SEQ ID NO: 49	496 bp	
NOV12g, 170534177 DNA Sequence	GGGTACCTGTGGCTATACCCAGGACCTGACAGACAACCTTGACTGGACGCGGCAGAAATGCCCTCAGCCAGAACCCCAACGCTCCCCCAACACTGGTCCCCCACCAGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCTCGGGAGCTGGGGGACCGTGCAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTTCTACTGTGTCTCTCTTCTATCACATGTACGGGAAACACATCGGCTCCCTCAACCTCCTGGTGCGGTCCC GGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCAATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGTCCCAGTGGGCCCCCTCCAGATTATTTTGTAGGGGGTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGACGTCACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGAATTC		
	ORF Start: GGT at 2	ORF Stop: at 497	
	SEQ ID NO: 50	165 aa	MW at 18420.5kD
NOV12g, 170534177 Protein Sequence	GTCGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELGDRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLVRSRNKGALDTHAWSLSGNKGNVWQAHVPI SPSPGFQII FEGVRGPGYLGDI AIDDVTLKKGECPRKQTEF		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b through NOV12g.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	239..536 3..300	283/298 (94%) 283/298 (94%)
NOV12c	239..536 3..300	284/298 (95%) 284/298 (95%)
NOV12d	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12e	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12f	683..959 3..239	234/277 (84%) 235/277 (84%)
NOV12g	801..962 3..164	161/162 (99%) 162/162 (99%)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

<b>Table 12C. Protein Sequence Properties NOV12a</b>	
<b>PSort analysis:</b>	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
<b>SignalP analysis:</b>	Likely cleavage site between residues 19 and 20

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 12D.

<b>Table 12D. Geneseq Results for NOV12a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV12a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE00582	Human nuclear cell adhesion molecule homologue, NCAM_d_1 protein - Homo sapiens, 946 aa. [WO200129215-A2, 26-APR-2001]	23..959 15..912	487/946 (51%) 656/946 (68%)	0.0
AAE00581	Human cell adhesion molecule homologue (CAM-H) protein #1 - Homo sapiens, 1018 aa. [WO200129215-A2, 26-APR-2001]	23..959 15..912	487/946 (51%) 656/946 (68%)	0.0
AAE00586	Human nuclear cell adhesion molecule homologue, NCAM_d_2 protein - Homo sapiens, 891 aa. [WO200129215-A2, 26-APR-2001]	71..959 8..857	455/898 (50%) 617/898 (68%)	0.0
AAAY72717	HBXDJ03 clone human attractin-like protein #2 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..965 1..418	416/458 (90%) 417/458 (90%)	0.0
AAAY72714	HBXDJ03 clone human attractin-like protein #1 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..965 1..418	406/458 (88%) 407/458 (88%)	0.0

- In a BLAST search of public sequence databases, the NOV12a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 12E.

<b>Table 12E. Public BLASTP Results for NOV12a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV12a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>

CAB86654	DJ402N21.3 (NOVEL PROTEIN WITH IMMUNOGLOBULIN DOMAINS) - Homo sapiens (Human), 299 aa (fragment).	239..536 1..299	298/299 (99%) 298/299 (99%)	e-172
CAB86653	DJ402N21.2 (NOVEL PROTEIN WITH MAM DOMAIN) - Homo sapiens (Human), 273 aa (fragment).	683..965 1..243	242/283 (85%) 242/283 (85%)	e-138
Q9DBX0	1200011I03RIK PROTEIN - Mus musculus (Mouse), 267 aa.	689..965 1..237	227/277 (81%) 232/277 (82%)	e-129
Q9GMT4	HYPOTHETICAL 51.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 448 aa.	508..959 1..414	205/461 (44%) 281/461 (60%)	e-109
CAB86655	DJ402N21.1 (NOVEL PROTEIN) - Homo sapiens (Human), 127 aa (fragment).	1..127 1..127	127/127 (100%) 127/127 (100%)	3e-68

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 7	53..110	14/61 (23%) 42/61 (69%)	2.5e-08
ig: domain 2 of 7	150..216	14/70 (20%) 51/70 (73%)	3.7e-09
ig: domain 3 of 7	255..310	18/58 (31%) 38/58 (66%)	2.4e-08
PKD: domain 1 of 1	239..327	22/100 (22%) 56/100 (56%)	7.3
ig: domain 4 of 7	350..417	15/69 (22%) 49/69 (71%)	6.3e-11
ig: domain 5 of 7	456..516	18/64 (28%) 46/64 (72%)	1.7e-08
ig: domain 6 of 7	553..617	16/66 (24%) 39/66 (59%)	0.00011
fn3: domain 1 of 1	643..733	20/93 (22%) 53/93 (57%)	0.98
ig: domain 7 of 7	801..875	7/78 (9%) 54/78 (69%)	37

MAM: domain 1 of 1	793..958	65/180 (36%) 132/180 (73%)	1.3e-52
--------------------	----------	-------------------------------	---------

**EXAMPLE 13.**

The NOV13 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis		
	SEQ ID NO: 51	4169 bp
NOV13a, CG57409-03 DNA Sequence	TCTTCGTCGCGCTCTCTCTCTCACCTCTCAGGGAAAGGGGGGACATAGGGGCGTCG CGGGGCCCCGGCGAATGCGCCCCCGCGCCTCTCGGGCTGCGCCGCTCGCGGGGAT GAAGCACCGGCCGTGAAGATGGAGGTGACCTGCCTTCTACTTCTGGCGCTGATCCCT TCCACTGCGGGGACAAGGAGTCTACGCTCCAGCCAGGCGCAGATCGTGCATGCGGG CCAGGCATGTGTGGTGAAGAGGACAATATCAGCGAGCGTGTCTACCCATCCGGGAG GGGGACACCCTCATGCTGCAGTGCCTTGTAAAGGGCACCCTCGACCCAGGTACGGT GGACCAAGACGGCAGGTAGCGCTCGGACAAGTTCAGGAGACATCGGTGTTCAACGA GACGCTGCGCATCGAGCGTATTGCACGCACGAGGCGGCCGCTACTACTGCAAGGCT GAGAACGGCGTGGGGGTGCCGGCATCAAGTCCATCCGCGTGGACGTGCAGTACCTGG ATGAGCCAATGCTGACGGTGCACAGACGGTGAGCGATGTGCGAGGCAACTTCTACCA GGAGAAGACGGTGTCTCTGCGTGTACTGTCAACTCCAACCCGCTGCCGCTTCATC TGGAAGCGGGGTTCCGATACCTATCCACAGCCAGGACAATGGGGTTGACATCTATG AGCCCCCTTACTACTCAGGGGAGACCAAGGTCTGAAGCTGAAGAACCTGCGGCCCCA GGACTATGCCAGTACACCTGCCAGGTGTCTGTGCGTAACGTGTGCGGCATCCAGAC AAGGCCATCACCTTCCGGCTCACCAACACCAGGCACCAGCCCTGAAGCTGTCTG TGAACGAACTCTGGTGGTGAACCTGGGGAGAATGTGACGGTGCAGTGTCTGCTGAC AGGCGGTGATCCCCCTCCCCAGCTGCAGTGGTCCCATGGGCCTGGCCCACTGCCCTG GGTGCTCTGGCCAGGGTGGCACCCTCAGCATCCCTTCAGTGCAGGCCCCGGGACTCTG GCTACTACAATGCACAGCCACCAACAATGTGGGCAACCTGCCAAGAAGACTGTCAA CCTGCTGGTGCGATCCATGAAGAAGCTACATTCCAGATCACTCTGACGTGATCAAA GAGAGTGAGAACATCCAGCTGGGCCAGGACCTGAAGCTATCGTGCCACGTGGATGCAG TGCCCCAGGAGAAGGTGACCTACCAGTGGTTCAAGAATGGCAAGCCGGCAGCATGTC CAAGCGGCTGCTGGTGACCGCAATGATCCTGAGCTGCCCGCAGTCACCAGCAGCCTA GAGCTCATTGACCTGCACTTCAGTGACTATGGCACCTACCTGTGCATGGCTTCTTTC CAGGGGCAACCGTGCCCGACCTCAGCGTCGAGGTCAACATCTCTGAGCATGCGCC GCCCACCATCAGTGTGCCAAGGGTAGGGCCGTGGTGACCGTGCAGGAGGATCGCCT GCCGAGCTGCAATGCGAGGTGCGGGCAAGCCGCGCCGCGCAGTGTCTGGTCCGCG TGGACAAGGAGGCTGCACTGCTGCCCTCGGGGCTGCCCTGGAGGAGACTCCGGACCG GAAGCTGCGGCTGGAGCGAGTGAGCCGAGACATGAGCGGGACCTACCGTGCCAGACG GCCCCTATAATGGCTTCAACGTGCGCCCCGTGAGGCCAGGTGCAGTGAACGTGC AGTTCGCCCGGAGGTGGAGCCAGTTCACAGGACGTGCCAGGCGCTGGGCCGGCC CGTGCTCCTGCGCTGCTCGTGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTG TGGCGTTTCAAAGGGCAGCTGCTGCCGCGCCGCGCTGTTGTTCCCGCCGCGCCGAGG CGCCGGATCAGCGGAGCTGCGCCTCGACGCGCTAACTCGCGACAGCAGCGGACGTA CGAGTGCAGCGTCTCAACGATGTGGGCTCGGCTGCCTGCTTCCAGGTCTCCGCC AAAGCCTACAGCCCGGAGTTTACTTCGACACCCCCAACCCACCCGAGCCACAAGC TGTCCAAGAACTACTCTACGTGCTGCAGTGGACTCAGAGGGAGCCGACGCTGTGCA CCCTGTGCTCAACTACAGACTCAGCATCCGCCAGTTGAACCAGCACAAATGGGTGGTC AAGGCCATCCCGGTCCGGCGTGTGGAGAAGGGGAGCTGCTGGAGTACATCCTGACCG ATCTCCGTGTGCCCCACAGCTATGAGGTCCGCTCACACCTATACCACCTTGGGGC TGGTGACATGGCTCCCGCATCATCCACTACACAGAGCCCATCAACTCTCCGAACCTT TCAGACAACACCTGCCACTTTGAGGATGAGAAGATCTGTGGCTATACCCAGGACCTGA CAGACAACCTTTGACTGGACGCGGCAGAAATGCCCTACCCAGAACCCAAACGCTCCCC CAACACTGGTCCCCCACCAGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTATC GAGACATCGAGGCTCGGGAGCTGGGGGACCGTGAAGGTTAGTGAGTCCCCTCTACA ATGCCAGCGCCAAGTTCTACTGTGTCTCTTCTTACCACATGTACGGGAAACACAT CGGCTCCCTCAACCTCCTGGTGGGTCCCGAACAAGGGGCTCTGGACACGCACGCC TGGTCTCTCAGTGGCAATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCC	



	CCAGTGGGCCCTTCCAGATTATTTTGGAGGGGTTGAGGCCCGGGCTACCTGGGGGA TATTGCCATAGATGACGTACACTGAAGAAGGGGAGTGTCCCCGAAGCAGACGGAT CCCAATAAAGGTGCAAGACGGGAAGGAGCTGCCTGCGATGGCCTGAAATTCCACCTTT CATCCCCCTATGGATGACGGAGAGCTTACAGATGACCCTATTGAATGCAAGCACCTTTG GATCCATAGAGTGGACAGTAAAGGTGCTCAGTACATGTTGGCTGAGCTGAACTGCATA CATGTGGCCCCCAGGTTCTCTGGTCTTTATGGACGAAGGGCACAAGGTTGGTGAAAAGG ACTCCGGGGGCCAGCCCTTCCAAGTTTACACTGATTTCCTCTTTTACCCTCATGCTAT CCCTGAGAAGATGTCAATAATGCCACGTTACAGGTGGGAAACTGAGGCTTAGAGAG GAGGAGGAATCTGCCTACGGTCACACAGCTGCAAAGGCTAGAGCTGGGACCAGGAGCT GGTCTCTTAACCGACCACCTGAGCTCAAGAGCTTTTCTCTCTGGACCAACATGACCC AAGTGTGCGCGAGCCTATCACAGGTCCCCTGCAATGCCAAACATACACGCACAGCAAT ACACAACACCTGGGGACATGGATGAAGCTGGAAACCATCATTTCTCAGCAAACCTGACAC AAGAACAGAAAAACCAACACCATGTTCTCACTCACCACCCAGTCTGCCCCGCCCTC TCTCTTCTCACCTGAACTTCCCCTCTCTCAAACCTCTCGAGGCCACGCCTCTATGTCC TTGGATGATGATGATGACGACGACGACGATGATGATGATGATGATGACGACGATGACA ATGATGATGATGATGGAAGGAAGACCTACAGAATCCCTCCAGGCTCTGACCTCAGTGC TTGTGGGTGGGTGAATGACCACATGTGCGAGGGAGACTCCACAGGTCTCTCCGATGAG AAGCACTCTTATGCCAAAGAGGAGACTCAGGCCAACTGACAGGACCAGGAATTAGCT ACCCTGGTAAACCCAGCTATCGACTGCACCCGAGCGGCTACACACCACTGGAGCAGTT CAGGGAGAAAGCCACCGGCATGCTCACCCCGTATGTCTCTGGCTCTGTTTCTCTTTT TGCTTCCCCTTCCCACCTCTGAGTCTCTGTGTCTGCTCATGCCAATTCCCCTCTG CCTGTCTCTGCCCGCTTCTCTCTGGGCTGGTCTCTCCGAGACTCTGTTCCCTTGGCTG GCATGCCCTCCACCTCCCCTGATGGTTCAGCAGAGATGAAGCCGCGCTGGCTCATGGG TGTGGTAAATGTACTAGTCAGGAGAGTGGTGGGGCCAGTCTGGGTGCAG		
	ORF Start: ATG at 135	ORF Stop: TGA at 4080	
	SEQ ID NO: 52	1315 aa	MW at 145782.9kD
NOV13a, CG57409-03 Protein Sequence	MEVTCLLLLALIPFHCRCQGVYAPAAQIVHAGQACVVKEDNISERVYTIREDTLM LQCLVTGHPRPQVRWTKTAGSASDKFQETSVFNELRIERIARTQGGRYCKAENGVG VPAIKSIRVDVQYLDEPMLTVHQTVSDVRGNFYQKTVFLRCTVNSNPPARFIWKRGS D TLSHSQDNGVDIYEPLYTQGETKVLKLNLRPDYASYTCQSVVRNVCIPDKAITFR LTNTTAPPALKLSVNETLVNPGENVTVQCLLTGGDPLPQLQWSHGPGPLPLGALAQ G GTLSIPSVQARDSGYYNCTATNNVGNPAKKTVNLLVRSMKNATFQITPDVIKES ENIQ LGQDLKLSCHVDVAPQEKVTYQWFKNGKPARMSKRLLVTRNDPELPAVTS SLELIDLH FSDYGTLYCMASFPGAPVPDLSVEVNISSETVPTTISVPKGRAVTV TREGSPAELQCE VRGKPRPPVLSVRVDKEAALLPSGLPLEETPDGKRLRLERSVR DMSGTYRCQTARYNGF NVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLLRCS LLRGSPQRIASAVWRFKQG LLPPPPVVPAAAEAPDHAELRLDAVTRDSSGSYEC SVSNDVGSAAFLQVSAKAYSPE FYFDTNPNTSRHKLKSNYSYVLQWTQREPD VDPVLNRYLSIRQLNQHNNAVVKAI PVR RVEKQQLLEYILTLDRVPHSYEVRLTPYTTFGAGDMASRIHYTEPINSP NLSDNSTCH FEDEKICGYTQDLTDNFDWTRQNALTONPKRSPNTGPPTDISGT PEGYMFIETSRPR ELGDRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNLLV RSRNGALDTHAWSLGN KGNVWQQAHPVIPSPPFQIIIFEGVRGPGYLGDIADDT VLKKGECPRKQTPDNKGAR REGAACDGLKFHLSSPMDDGELTDDPIECKHLW IHRVDSKGAQYMLAELNCIHVAPRF LVFMDEGHKVGKDSGGQPFQVYTD DFSYPHAIPEKMSIMPTLQVGKRLRREEESAY GHATAKARAGTRSWSLN RPPELKSFSLWTNMTQSVREPITGPLQCQTYTHSNTQHLGT WMKLETIILSK LTQEQTKHHMFSLTQSAAPSLFSPPELPLSSNSRGHASMSLDDDDDD DDDDDDDDDDDDDDDDGRKTYRIPPGSDLSACGWVNDHMSQGDSTGPPDEK HSYAK EETQAKLTGPGISYPGKPSYRLHPSGYTFLEQFREKATGMLTPYVSG SVSSFCFPFPT SESLCSAHANSPSACLCPLLSGLVSPRLCSLGHWHALHP		
	SEQ ID NO: 53	1500 bp	
NOV13b, CG57409-05 DNA Sequence	TGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAA CGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGCAGTTCCCGCCGAGGTGGAG CCCAGTCCCAGGACGTGCCCGAGGCGCTGGGCGGCCCCGTGCTCCTGCGCTGCTCGC TGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTGTGGCGTTTCAAAGGGCAGCT GCTGCCGCGCGCGCCTGTTGTTCCCGCCGCGCCGAGGCGCCGGATCACGCGGAGCTG CGCCTCGACGCCGTAACCTCGCAGCAGCAGCGGACGCTACGAGTGCAGCGCTCCAAACG ATGTGGGCTCGGCTGCCTGCTCTTCCAGGTCTCCGCCAAAGCCTACAGCCCGAGTT TTACTTCGACACCCCCAACCCACCCGAGCCACAAGCTGTCCAAGAACTACTCTCAAC		

	GTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTGCACCTGTGCTCAACTACAGAC TCAGCATCCGCCAGTTGAACCAGCACAATGCGGTGGTCAAGGCCATCCCGGTCCGGCG TGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGC TATGAGGTCCGCCTCACACCTATACACCTTCGGGGCTGGTGACATGGCCTCCCGCA TCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCAGTCTCTGGCTCTGAGGACCT GTCTCTGGTCCCAAGCAGGGTATCCTCTGCAGAGCCCCAACCTCAGTTCTGACTTG GTTTCCCCGCTTGCTTTCTCAGCCATCAACTCTCCGAACCTTCAGACAACACCTGCC ACTTTGAGGATGAGAAGATCTGTGGCTATAACCAGGACCTGACAGACAACCTTGACTG GACGCGGCAGAAATGCCCTCACCAGAACCCCAAACGCTCCCCAACACTGGTCCCCC ACCGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCT GGGAGCTGGGGGACCGTGAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTT CTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCCC CTGGTGCAGTCCCGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCA ATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCA GATTATTTTGGAGGGGTTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGAC GTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGATCCCAATAAAGTGGTGG TGATGCCGGGCAGTGGAGCCCCCTGCCAGTCCAGCCCACAGCTGTGGGGGCCCATGGC CATCTTCTCTTGCGGTTGCAGAGATGATGAGAGCTGTGTGGCCACCCC		
	ORF Start: ATG at 12	ORF Stop: TGA at 1476	
	SEQ ID NO: 54	488 aa	MW at 54357.1kD
NOV13b, CG57409-05 Protein Sequence	MSGTYRCQTARYNGFNVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLLRCSLLRG SPQRIASAVWRFKGQLPPPPVPPAAEAPDHAELRLDAVTRDSSGSYECVSNDVGS AACLFQVSAKAYSPEFYFDTPNPTRSHKLSKNYSYVLQWTQREPDAVDPVLYRLSIR QLNQHNNAVVKAIIPVRRVEKGQLLEYILTLRVPHSYEVRLTPYTTFGAGDMASRIHY TERQIRWPPVLALRTLSSGPKQGILCRAPHLSSDLVSPPLAFSAINSPNLSNTHCFED EKICGYTQDLTDNFDWTRQNALTONPKRSPNTGPPTDISGTPEGYYMFIETSRPRELG DRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNPLVRSRNGALDTHAWSLSGNKGN VWQQAHPVISPSPGFQIIFEGVRGPGYLGDIADDDVTLKKGECPRKQTDPNKVVMPG SGAPCQSSPQLWGPMAIFLLALQR		
	SEQ ID NO: 55	1828 bp	
NOV13c, CG57409-06 DNA Sequence	TGAGCCGAGACATGAGCGGGACCTACCGCTGCCAGACGGCCCGCTATAATGGCTTCAA CGTGCGCCCCCGTGAGGCCAGGTGCAGCTGAACGTGCAGTTCCCGCGGAGGTGGAG CCCAGTTCCAGGACGTGCCAGGCGCTGGGCGGGCCCGTGCTCCTGCGCTGCTCGC TGCTGCGAGGCAGCCCCAGCGCATCGCCTCGGCTGTGTGGCGTTTCAAAGGGCAGCT GCTGCCGCCCGCGCTGTTGTTCCCGCGCGCGCGAGGCGCGGATCAGCGGAGCTG CGCCTCGACGCCGTAACCTCGCGACAGCAGCGGCAGCTACGAGTGCAGCGTCTCCAACG ATGTGGGCTCGGCTGCCTGCCTCTTCCAGGTCTCCGCCAAAGCCTACAGCCCGGAGTT TTACTTCGACACCCCCAACCCACCCGAGCCACAAGCTGTCCAAGAATACTCCTAC GTGCTGCAGTGGACTCAGAGGGAGCCCGACGCTGTCGACCCTGTGCTCAACTACAGAC TCAGCATCCGCCAGTTGAACCAGCACAATGCGGTGGTCAAGGCCATCCGGTCCGGCG TGTGGAGAAGGGGCAGCTGCTGGAGTACATCCTGACCGATCTCCGTGTGCCCCACAGC TATGAGGTCCGCCTCACACCTATACACCTTCGGGGCTGGTGACATGGCCTCCCGCA TCATCCACTACACAGAGCGCCAGATCCGCTGGCCCCCAGTCTCTGGCTCTGAGGACCT GTCTCTGGTCCCAAGCAGGGTATCCTCTGCAGAGCCCCACACCTCAGTTCTGACTTG GTTTCCCCGCTTGCTTTCTCAGCCATCAACTCTCCGAACCTTCAGACAACACCTGCC ACTTTGAGGATGAGAAGATCTGTGGCTATACCCAGGACCTGACAGACAACCTTGACTG GACGCGGCAGAAATGCCCTCACCAGAACCCCAAACGCTCCCCAACACTGGTCCCCC ACCGACATAAGTGGCACCCCTGAGGGCTACTACATGTTTCATCGAGACATCGAGGCCT GGGAGCTGGGGGACCGTGAAGGTTAGTGAGTCCCCTCTACAATGCCAGCGCCAAGTT CTACTGTGTCTCCTTCTTCTACCACATGTACGGGAAACACATCGGCTCCCTCAACCCC CTGGTGCAGTCCCGGAACAAAGGGGCTCTGGACACGCACGCCTGGTCTCTCAGTGGCA ATAAGGGCAATGTGTGGCAGCAGGCCCATGTGCCCATCAGCCCCAGTGGGCCCTTCCA GATTATTTTGGAGGGGTTTCGAGGCCCGGGCTACCTGGGGGATATTGCCATAGATGAC GTCACACTGAAGAAGGGGGAGTGTCCCCGGAAGCAGACGGATCCCAATAAAGTGCAA GACGGGAAGGAGGTGGGGGAGCTGAATCTGGAGGGAGCTGTGCGTGGCGGGGGTTCTT GTCTGTTGAGGGAGGGTGTTCGGGTCTGAATAGGGGTTCAAGTGTCTGATGATGGGA ATCAGGTGGCTCTGACTGTGTTAACGTGTGCCCAACTCAGCTCAGGCTGAGAAGTGT GTGTAACACCATGAGAAAGCTTGGCCCCACCATCGTGATGAGCATAACGACCTGGTC		

	ACCGGAACACAAACACCAACACACAGAGGGCGCCTCAGAATACCCAGAGGGCCCAA TACGCCGACCCGCTGTACAGAGCGCCACGAGCGGCAGAACACGACAGGCACACAACC AGCCGGAGCAAGACGGAGCCGAGAGCCCGGGGACATAGACCCAGCAAGCGACACAC AAGGACGCGCACAGAGCGCACACTAACA		
	ORF Start: ATG at 12	ORF Stop: TGA at 1521	
	SEQ ID NO: 56	503 aa	MW at 55764.4kD
NOV13c, CG57409-06 Protein Sequence	MSGTYRCQTARYNGFNVRPREAQVQLNVQFPPEVEPSSQDVRQALGRPVLRLRCSLLRG SPQRIASAVWRFGQLLPPPPVPAAEAPDHAE LR L DAVTRDSSGSYECSVSNDVGS AACL FQVSAKAYSPEFYFDTPNPT RSHKLSKNYSYVLQWTQREPDADVPLNYRLSIR QLNQHN AVVKAIPVRRVEKGQLLEYILTLRVPHSYEVRLTPYTTFGAGDMASRIIHY TERQIRWPPVLALRTLSSGPKQGILCRAPHLSSDLVSPLAFSAINSPNLS DNTCH FED EKICGYTQDLTDNFDWTRQNALTQNPKRSPNTGPPTDISGTPEGYYMFIETSRPRELG DRARLVSPLYNASAKFYCVSFFYHMYGKHIGSLNPLVRSRNGALDTHAWSLSGNKGN VWQQAHVPISPSGPFQIIIFEGVRGPGYLGDI AID DVTLKKGECPRKQTDPNKGARREG GGGAESGGSCAWRGFLSVEGGCSGLNRGSDCLMMGIRWL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 13B.

Table 13B. Comparison of NOV13a against NOV13b through NOV13c.		
Protein Sequence	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV13b	508..925 1..458	403/458 (87%) 403/458 (87%)
NOV13c	508..925 1..458	403/458 (87%) 403/458 (87%)

Further analysis of the NOV13a protein yielded the following properties shown in Table 13C.

Table 13C. Protein Sequence Properties NOV13a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

- 5 A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 13D.

Table 13D. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAE00582	Human nuclear cell adhesion molecule homologue, NCAM_d_1 protein - Homo sapiens, 946 aa. [WO200129215-A2, 26-APR-2001]	23..919 15..912	488/906 (53%) 657/906 (71%)	0.0
AAE00581	Human cell adhesion molecule homologue (CAM-H) protein #1 - Homo sapiens, 1018 aa. [WO200129215-A2, 26-APR-2001]	23..919 15..912	488/906 (53%) 657/906 (71%)	0.0
AAE00586	Human nuclear cell adhesion molecule homologue, NCAM_d_2 protein - Homo sapiens, 891 aa. [WO200129215-A2, 26-APR-2001]	71..919 8..857	456/858 (53%) 618/858 (71%)	0.0
AA72717	HBXDJ03 clone human attractin-like protein #2 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..925 1..418	418/418 (100%) 418/418 (100%)	0.0
AA72714	HBXDJ03 clone human attractin-like protein #1 - Homo sapiens, 448 aa. [WO200116156-A1, 08-MAR-2001]	508..925 1..418	408/418 (97%) 408/418 (97%)	0.0

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13E.

Table 13E. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAB86654	DJ402N21.3 (NOVEL PROTEIN WITH IMMUNOGLOBULIN DOMAINS) - Homo sapiens (Human), 299 aa (fragment).	239..536 1..299	298/299 (99%) 298/299 (99%)	e-172
CAB86653	DJ402N21.2 (NOVEL PROTEIN WITH MAM DOMAIN) - Homo sapiens (Human), 273 aa (fragment).	683..925 1..243	243/243 (100%) 243/243 (100%)	e-145
Q9DBX0	1200011I03RIK PROTEIN - Mus musculus (Mouse), 267 aa.	689..925 1..237	228/237 (96%) 233/237 (98%)	e-136
Q9GMT4	HYPOTHETICAL 51.2 KDA PROTEIN - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 448 aa.	508..919 1..414	206/421 (48%) 282/421 (66%)	e-115
CAB86655	DJ402N21.1 (NOVEL PROTEIN) - Homo sapiens (Human), 127 aa (fragment).	1..127 1..127	127/127 (100%) 127/127 (100%)	3e-68

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13F.

Table 13F. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 7	53..110	14/61 (23%) 42/61 (69%)	2.5e-08
ig: domain 2 of 7	150..216	14/70 (20%) 51/70 (73%)	3.7e-09
ig: domain 3 of 7	255..310	18/58 (31%) 38/58 (66%)	2.4e-08
PKD: domain 1 of 1	239..327	22/100 (22%) 56/100 (56%)	7.3
ig: domain 4 of 7	350..417	15/69 (22%) 49/69 (71%)	6.3e-11
ig: domain 5 of 7	456..516	18/64 (28%) 46/64 (72%)	1.7e-08
ig: domain 6 of 7	553..617	16/66 (24%) 39/66 (59%)	0.00011
fn3: domain 1 of 1	643..733	20/93 (22%) 53/93 (57%)	0.98
ig: domain 7 of 7	761..835	7/78 (9%) 54/78 (69%)	37
MAM: domain 1 of 1	753..918	65/180 (36%) 132/180 (73%)	1.3e-52

**EXAMPLE 14.**

The NOV14 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 57		330 bp
NOV14a, CG59262-01 DNA Sequence	GGAGTGGTCAGTTCTGCTGCCGACACGCCACCCAGCTCGAGATGGCCATGGACACCA TGATTAGAATCTTCCACCGCTATTCTGGCAAGGCAAGGAAGAGATTCAAGCTCAGCA GGGGGAAGTGAAGTCTCCTGCAGCGAGAGCTCAGGAATTCCTCTCGTGCCAAAAG GAAACCCAGTTGGTTGATAAGATAGTGCAGGACCTGGATGCCAATAAGGACAACGAAC TGGATTTTAATGAATTCTGGTTCATGGTGGCAGCTCTGACAGTTGCTTGTAATGATT CTTTGTAGAACAATTGAAGAAGAAAGGAAAATAAAGGTAA		
	ORF Start: ATG at 43		ORF Stop: TAA at 322
	SEQ ID NO: 58		93 aa      MW at 10861.6kD

NOV14a, CG59262-01 Protein Sequence	MAMDTMIRIFHRYSGKARKRFLSKGELKLLQLRELTEFLSCQKETQLVDKIVQDLDA NKDNEVDFNEFVVMVAALTVCNDYFVEQLKKKGK
---	---

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.5337 probability located in mitochondrial inner membrane; 0.3627 probability located in mitochondrial intermembrane space; 0.2997 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 14C.

Table 14C. Geneseq Results for NOV14a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM40258	Human polypeptide SEQ ID NO 3403 - Homo sapiens, 94 aa. [WO200153312-A1, 26-JUL-2001]	2..86 8..92	50/85 (58%) 66/85 (76%)	3e-23
AAB45531	Human S100A1 protein - Homo sapiens, 94 aa. [DE19915485-A1, 19-OCT-2000]	2..86 8..92	50/85 (58%) 66/85 (76%)	3e-23
ABB12007	Human Ca-binding protein S100P homologue, SEQ ID NO:2377 - Homo sapiens, 113 aa. [WO200157188-A2, 09-AUG-2001]	2..84 25..107	43/83 (51%) 59/83 (70%)	3e-18
AAB45545	Human S100P protein - Homo sapiens, 95 aa. [DE19915485-A1, 19-OCT-2000]	2..84 7..89	43/83 (51%) 59/83 (70%)	3e-18
AAB45544	Human S100B protein - Homo sapiens, 95 aa. [DE19915485-A1, 19-OCT-2000]	2..84 7..89	43/83 (51%) 59/83 (70%)	3e-18

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. Public BLASTP Results for NOV14a				
Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAL30893	S100Z PROTEIN - Homo sapiens (Human), 99 aa.	1..93 7..99	93/93 (100%) 93/93 (100%)	4e-47
S35985	S-100 protein alpha chain - weatherfish, 95 aa.	2..89 7..94	52/88 (59%) 70/88 (79%)	3e-25
P35467	S-100 protein, alpha chain - Rattus norvegicus (Rat), 93 aa.	2..86 7..91	52/85 (61%) 66/85 (77%)	4e-23
BCBOIA	S-100 protein alpha chain - bovine, 94 aa.	2..86 8..92	50/85 (58%) 66/85 (76%)	1e-22
CAC16547	SEQUENCE 1 FROM PATENT WO0061742 - Homo sapiens (Human), 94 aa.	2..86 8..92	50/85 (58%) 66/85 (76%)	1e-22

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14E.

Table 14E. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
S_100: domain 1 of 1	2..42	20/44 (45%) 31/44 (70%)	2.8e-09
efhand: domain 1 of 1	48..76	6/29 (21%) 25/29 (86%)	0.0012

#### EXAMPLE 15.

The NOV15 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis		
	SEQ ID NO: 59	773 bp
NOV15a, CG58635-01 DNA Sequence	AGCCTTGGGTCGAAGGGATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGT CCGTGGGTTCTCTGTCCCTACAGAAAAGGAGAACAACTTCCCGCCACTGCCCAAGTTC ATCCCTGTGAAGCCCTGCTTCTACCAGAACTTCTCCGACGAGATCCAGTGAGCACC AGGTCTGGTGAAGAGGATCTACCGGCTGTGGATGGTTTACTGCGCCACCCTCGGCGT CAACCTCATTGCCTGCCTGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTCGGC CTGGCCTTCGTGTGGCTGCTCCTGTTACGCCTTGCGGCTACGTGTGCTGGTTCCGGC CTGTCTACAAGGCCTTCCGGGCCGACAGCTCCTTTAATTTATGGCGTTTTTCTTCAT CTTCCGAGCCCAGTTTGTCTGACCGTCATCCAGGCGATTGGCTTCTCCGGCTGGGGC GCGTGCGGCTGGCTGTCCGCAATTGGATTCTTCCAGTACAGCCCGGGCGCTGCCGTGG TCATGCTGCTTCCAGCCATCATGTTCTCCGTGTGCGCTGCCATGATGGCCATCGCGAT	

	CATGAAGGTGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACG GAGTGGAAACACGGGCACCTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAAC TCTCAGGCAACAGCCTGCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCA GTGGCCTTAGAGGGAGCCT		
	ORF Start: ATG at 18	ORF Stop: TAG at 762	
	SEQ ID NO: 60	248 aa	MW at 27780.0kD
NOV15a, CG58635-01 Protein Sequence	MRWGLLQETKSGSVRGFSVPTEKENNFPLPKFIPVKPCFYQNFSDIIPVEHQVLVKR IYRLWMVYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLFTPCGYVCWFRPVYKAF RADSSFNMAFFFI FGAQFVLTVIQAI GFSGWACGWLSAIGFFQYSPGAAVMMLPA IMFSVSAAMMAIAIMKVHRIYRGAGGSFQKAQTEWNTGTWRNPPSREAQYNNFSGNSL PEYPTVPSYPGSGQWP		
	SEQ ID NO: 61	773 bp	
NOV15b, CG58635-02 DNA Sequence	AGCCTTGGGTTGAAGGGATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGT CCGTGGGTTCTCTGTCCCTACAGAAAAGGAGAACAACCTCCCGCCACTGCCCAAGTTC ATCCCTGTGAAGCCCTGCTTCTACCAGAACTTCTCCGACGAGATCCCACTGGAGCACC AGGTCTGGTGAAGAGGATCTACCGGCTGTGGATGTTTTACTGCGCCACCCTCGGCGT CAACCTCATTGCCTGCCTGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTTCGGC CTGGCCTTCGTGTGGCTGCTCCTGTTACGCGCTTGGGCTACGTGTGCTGGTTCGGGC CTGTCTACAAGGCCTTCGAGCCGACAGCTCCTTTAATTTATGGCGTTTTCTTCAT CTTCGGAGCCCAAGTTTGTCTGACCGTCATCCAGGCGATTGGCTTCTCCGGCTGGGGC GCGTGGCGCTGGCTGTGCGCAATTGGATTCTTCCAGTACAGCCCGGCGCTGCCGTGG TCATGCTGCTTCCAGCCATCATGTTCTCCGTGTGCGGTGCCATGATGGCCATCGCGAT CATGAAGGTGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACG GAGTGGAAACACGGGCACCTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAAC TCTCAGGCAACAGCCTGCCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCA GTGGCCTTAGAGGGAGCCT		
	ORF Start: ATG at 18	ORF Stop: TAG at 762	
	SEQ ID NO: 62	248 aa	MW at 27828.1kD
NOV15b, CG58635-02 Protein Sequence	MRWGLLQETKSGSVRGFSVPTEKENNFPLPKFIPVKPCFYQNFSDIIPVEHQVLVKR IYRLWMFYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLFTPCGYVCWFRPVYKAF RADSSFNMAFFFI FGAQFVLTVIQAI GFSGWACGWLSAIGFFQYSPGAAVMMLPA IMFSVSAAMMAIAIMKVHRIYRGAGGSFQKAQTEWNTGTWRNPPSREAQYNNFSGNSL PEYPTVPSYPGSGQWP		
	SEQ ID NO: 63	654 bp	
NOV15c, CG58635-03 DNA Sequence	ATGAGGTGGGGCCTCCTTCAAGAGACAAAGTCTGGTTCTGTCCGTGGGTTCCCGGTCC CTACAGAAAAGGAGAACAACCTCCCGCCACTGCCCAAGTTCATCCCTGTGAAGCCCTG CTTCTACCAGAACTTCTCCGACGAGATCCCACTGGAGCACCAGGTCTGGTGAAGAGG ATCTACCGGCTGTGGATGTTTTACTGCGCCACCCTCGGCGTCAACCTCATTGCCTGCC TGGCCTGGTGGATCGGCGGAGGCTCGGGGACCAACTTCGGCCTGGCCTTCGTGTGGCT GCTCCTGTTACGCGCTTCGGCTACGTGTGCTGGTTCCGGCCTGTCTACAAGGCCTTC CGCGGCTGGCTGTGCGCAATTGGATTCTTCCAGTACAGCCCGGCGCTGCCGTGGTCA TGCTGCTTCCAGCCATCATGTTCTCCGTGTGCGCTGCCATGATGGCCATCGCGATCAT GAAGGCGCACAGGATCTACCGAGGGGCTGGCGGAAGCTTCCAGAAGGCACAGACGGAG TGGAACACGGGCACTTGGCGGAACCCACCGTCGAGGGAGGCCAGTACAACAACCTTCT CAGGCAACAGCCTGCCCGAGTACCCCACTGTGCCAGCTACCCGGGCAGTGGCCAGTG GCCTTAGAGGGAGCCT		
	ORF Start: ATG at 1	ORF Stop: TAG at 643	
	SEQ ID NO: 64	214 aa	MW at 24129.8kD
NOV15c, CG58635-03 Protein Sequence	MRWGLLQETKSGSVRGFPVPEKENNFPLPKFIPVKPCFYQNFSDIIPVEHQVLVKR IYRLWMFYCATLGVNLIACLAWWIGGSGTNFGLAFVWLLFTPCGYVCWFRPVYKAF RWLSAIGFFQYSPGAAVMMLPAIMFSVSAAMMAIAIMKAHRIYRGAGGSFQKAQTE WNTGTWRNPPSREAQYNNFSGNSLPEYPTVPSYPGSGQWP		



Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

<b>Table 15B. Comparison of NOV15a against NOV15b through NOV15c.</b>		
<b>Protein Sequence</b>	<b>NOV15a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV15b	1..248	217/248 (87%)
	1..248	217/248 (87%)
NOV15c	1..248	191/248 (77%)
	1..214	191/248 (77%)

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

<b>Table 15C. Protein Sequence Properties NOV15a</b>	
<b>PSort analysis:</b>	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
<b>SignalP analysis:</b>	Likely cleavage site between residues 17 and 18

- 5 A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 15D.

<b>Table 15D. Geneseq Results for NOV15a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV15a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAM93439	Human polypeptide, SEQ ID NO: 3078 - Homo sapiens, 229 aa. [EP1130094-A2, 05-SEP-2001]	21..248 2..229	226/228 (99%) 227/228 (99%)	e-138
AAM93704	Human polypeptide, SEQ ID NO: 3635 - Homo sapiens, 132 aa. [EP1130094-A2, 05-SEP-2001]	21..150 2..131	127/130 (97%) 129/130 (98%)	1e-75
AAM25225	Human protein sequence SEQ ID NO:740 - Homo sapiens, 185 aa. [WO200153455-A2, 26-JUL-2001]	21..131 35..145	109/111 (98%) 110/111 (98%)	1e-64
AAY11904	Human 5' EST secreted protein SEQ ID No: 504 - Homo sapiens, 108 aa. [WO9906550-A2, 11-FEB-1999]	21..126 2..107	102/106 (96%) 103/106 (96%)	7e-60

AAB62698	Human membrane recycling protein (HMRP)-1 - Homo sapiens, 347 aa. [US6235715-B1, 22-MAY-2001]	23..229 131..338	102/208 (49%) 140/208 (67%)	5e-56
----------	---	---------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a				
Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q969E2	HYPOTHETICAL 25.7 KDA PROTEIN (SIMILAR TO SECRETORY CARRIER MEMBRANE PROTEIN 4) - Homo sapiens (Human), 229 aa.	21..248 2..229	226/228 (99%) 227/228 (99%)	e-138
Q9ET20	SECRETORY CARRIER MEMBRANE PROTEIN 4 - Rattus norvegicus (Rat), 230 aa.	23..248 4..230	193/227 (85%) 208/227 (91%)	e-118
Q9JKV5	SECRETORY CARRIER MEMBRANE PROTEIN 4 - Mus musculus (Mouse), 230 aa.	23..248 4..230	190/227 (83%) 208/227 (90%)	e-117
Q9JKE3	SECRETORY CARRIER MEMBRANE PROTEIN 5 - Rattus norvegicus (Rat), 235 aa.	22..246 3..234	135/232 (58%) 167/232 (71%)	2e-81
Q9JKD3	SECRETORY CARRIER MEMBRANE PROTEIN 5 - Mus musculus (Mouse), 235 aa.	22..246 3..234	134/232 (57%) 166/232 (70%)	7e-81

Pfam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a			
Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TspO_MBR: domain 1 of 1	63..190	30/164 (18%) 91/164 (55%)	9.5
chloroa_b-bind: domain 1 of 1	181..195	5/15 (33%) 12/15 (80%)	3.7

**EXAMPLE 16.**

The NOV16 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 65	1642 bp	
NOV16a, CG59209-01 DNA Sequence	GCGGAGTCCGGACGTGCGGAGCAGGATGGCGGCGGAGCAGGACCCCGAGGCGCGCGCG GGGCGGCCGCTGCTCACTGACCTCTACCAGGCCACCATGGCGTTGGGCTATTGGCGCG CGGGCCGGGCGCGGACGCCGCCGAGTTCGAGCTCTTCTTCCGCCGTGCCCGTTCGG CGGCGCCTTCGCTTGGTAGCCGGCTTGCGCGACTGTGTGCGCTTCCTGCGCGCCTTC GACGTGCAGTTCTTGGCTCGGTGCTGCCCCAGACACGGATCCTGCGTTCTTCGAGC ACCTTCGGGCCCTCGACTGCTCCGAGGTGACGGTGCGAGCCCTGCCCGAGGCTCCCTC GCCTTCCCCGCGAGGTGCCGCTCCTGCAGGTGTCCGGGCCGCTCCTGGTGGTGCAGCTG CTGGAGACACCGCTGCTCTGCCTGGTCAGCTACGCCAGCCTGGTGGCCACCAACGCAG CGCGCGTTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAGAGATGGGCTGAGGCC GGCTCAGGGCCCGATGGGGCTGACAGCCTCCACCTACAGCTACCTGGGGGCTTCGAG AGCAGCAGCAACGTGCTAGCGGGCCAGCTGCGAGGTGTGCCGGTGGCCGGGACCTTGG CCCCTCCTTCGTCACTTCCCTTTTCAGGCAGCGAGGTGCCCCCTGACCCGATGTTGGC GCCAGCAGCTGGTGAGGGCCCTGGGGTGGACCTGGCGGCCAAAGCCAGGTGTGGCTG GAGCAGGTGTGTGCCACCTGGGGCTGGGGGTGACAGGAGCCGCATCCAGGCGAGCGGG CAGCCTTTGTGGCTATGCCTTGGCTTTTCCCGGGCCTTCAGGGCCTCCTGGACAC CTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCTAGCAGTCGCTTGGCCCTGGGA GAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGTGGTGACCTGCTACAGCAGG CTCAGGAGATCCGAAGGTCTTCCGAGCTGCTGCAGCCAGTTCAGAGTGCCTTGGCT GGAGTCAGTCTCATCGTAGTCAGCAACAACATTGACGAGGAGGCGCTGGCCCCACTG GCCCAGGAGGGCAGTGAGGTGAATGTATTGGCATTGGCACCAGTGTGGTCACTGCC CCCAACAGCCTTCCCTGGGTGGTGTCTATAAGCTGGTGGCCGTGGGGGGCCAGCCAC AATGAAGCTGACCGAGGACCCCGAGAAGCAGACGTTGCCTGGGAGCAAGGCTGCTTTC CGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATGCTGCAGTTAGCAGAAGAG CAGTGCCACAGGCTGGGCAGGAGCTGAGGGTGTGGCCTCCAGGGGCCAGGAGCCCTG CACCGTGAGGCCAGCCAGGTGGAGCCACTACTGCGGCTCTGCCTCCAGCAGGGACAG CTGTGTGAGCCGCTCCCATCTCCTGGCAGAGCTAGAGCCTTGGCCAGCTGTCCCTGA GCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTGCACAGTACCAGGTGGTGT GTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGCGGGGCAGTCCCCCTGAGAC TCGGAGCGGGGCTGACTG		
	ORF Start: ATG at 26	ORF Stop: TGA at 1619	
	SEQ ID NO: 66	531 aa	MW at 56889.8kD
NOV16a, CG59209-01 Protein Sequence	MAAEQDPEARAGRPLLTDLYQATMALGYWRAGRARDAAEFELFFRRCPFGGAFALVAG LRDCVRFLRAFVDVQFLASVLPDTPDAFFFEHLRALDCSEVTVRALPEAPSPSPQVPLL QVSGPLLVQLLETPLLCLVSYASLVATNAARVRLIAGPEKRLLMGLRRAQGPMPGD SLHLQLPGGFDSSSNVLQQLRGVPVAGTLAHSFVTSFSGSEVPPDPMLAPAAGEGPG VDLAAKAQVWLEQVCAHLGLGVQEPHPGERAAFVAYALAFPRAFQGLLDITYSVWRSGL PNFLAVALALGELGYRAVGVRDLSGDLQQAQEIIRKVFRAAAQFQVPWLESVLIVVS NNIDEALARLAQEGSEVNVIGIGTSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPE KQTLPGSKAAFRLLGSDGSPMLDMLQLAEEPVPQAGQELRVWPPGAQEPCTVRPAQVE PLLRLCLQQGQLCEPLPSLAESRALAQLSLRSLSPHRRRLSPAQYQVVLSERLQALV NSLCAGQSP		
	SEQ ID NO: 67	1179 bp	
NOV16b, 174308417 DNA Sequence	AGATCTACCAACGCAGCGCGCTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAG AGATGGGCTGAGGCGGGCTCAGGGCCCCGATGGGGGCTGACAGCCTCCACCTACAG CTACCTGGGCGGCTTCGACAGCAGCAGCAACGTGCTAGCGGGCCAGCTGCCAGGTGTG CCGGTGGCCGGGACCCTGGCCCACTCCTTCGTCACTTCTTTTCAGGCAGCAGGTGCG CCCCTGACCCGATGTTGGCGCCAGCAGCTGGTGAGGGCCCTGGGGTGGACCTGGCGGC CAAAGCCAGGTGTGGCTGGAGCAGGTGTGTGCCACCTGGGGCTGGGGGTGCAGGAG		

	CCGCATCCAGGCGAGCGGGCAGCCTTTGTGGCCTATGCCTTGGCTTTTCCCCGGGCCT TCCAGGGCCTCCTGGACACCTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCTTAGC AGTCGCCTTGGCCCTGGGAGAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGT GGTGACCTGCTACAGCAGGCTCAGGAGATCCGCAAGGTCTTCCGAGCTGCTGCAGCCC AGTTCAGGTGCCCTGGCTGGAGTCAGTCCTCATCGTAGTCAGCAACAACATTGACGA GGAGGCGCTTGGCCCGACTGGCCAGGAGGGCAGTGAGGTGAATGTCTATTGGCATTGGC ACCACTGTGGTCACCTGCCCCAACAGCCTTCCCTGGGTGGCGTCTATAAGCTGGTGG CCGTGGGGGGCCAGCCACGAATGAAGCTGACCGAGGACCCCGAGAAGCAGACGCTGCC TGGGAGCAAGGCTGCTTTCCGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATG CTGCAGTTAGCAGAAGAGCCAGTGCCACAGGCTGGGCAGGAGCTGAGGGTGTGGCCTC CAGGGGGCCAGGAGCCCTGCACCGTGAGGCCAGCCAGGTGGAGCCACTACTGCGGCT CTGCCTCCAGCAGGGACAGCTGTGTGAGCCGCTCCCATCCCTGGCAGAGTCTAGAGCC TTGGCCCAGCTGTCCCTGAGCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTG CACAGTACCAGGTGGTGTGTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGC GGGGCAGTCCCCCTCGAG		
	ORF Start: AGA at 1	ORF Stop: at 1180	
	SEQ ID NO: 68	393 aa	MW at 41797.4kD
NOV16b, 174308417 Protein Sequence	RSTNAARVRLIAGPEKRLLEMLRRAQGPDGGLTASTYSYLGFDSSSNVLAGQLRGV PVAGTLAHSFVTSFSGSEVPDPMLAPAAGEGPGVDLAAKAQVWLEQVCAHLGLGVQE PHPGERAAAFVAYALAFPRAFQGLLDITYSVWRSGLPNFLAVALALGELGYRAVGVRDLS GDLLQQAQEIIRKVFRAAAAFQVPWLESVLIVSNNIDEEALARLAQEGSEVNVIGIG TSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPEKQTLPGSKAAFRLLGSDGSPMDM LQLAEEPVPQAGQELRVWPPGAQEPCTVRPAQVEPLLRCLCQQQLCEPLPSLAESRA LAQLSLRSLSPHRRRLRSPAQYQVVLSERLQALVNSLCAGQSPLE		
	SEQ ID NO: 69	1179 bp	
NOV16c, 174308429 DNA Sequence	AGATCTACCAACGCAGCGCGCTTCGCTTGATCGCAGGGCCAGAGAAGCGGCTGCTAG AGATGGGCCTGAGCGGGCTCAGGGCCCCGATGGGGCCCTGACAGCCTCCACCTACAG CTACCTGGGCGGCTTCGACAGCAGCAGCAACGTGCTAGCGGGCCAGCTGCGAGGTGTG CCGGTGGCCGGGACCTGGCCCACTCCTTCGTCACTTCCTTTTCAGGCAGCGAGGTGC CCCCTGACCCGATGTTGGCGCCAGCAGCTGGTGAGGGCCCTGGGGTGGACCTGGCGGC CAAAGCCCAGGTGTGGCTGGAGCAGGTGTGTGCCACCTGGGGCTGGGGTGCAGGAG CCGCATCCAGGCGAGCGGGCAGCCTTTGTGGCCTATGCCTTGGCTTTTCCCCGGGCCT TCCAGGGCCTCCTGGACACCTACAGCGTGTGGAGGAGTGGTCTCCCCAACTTCTTAGC AGTCGCCTTGGCCCTGGGAGAGCTGGGCTACCGGGCAGTGGGCGTGAGGCTGGACAGT GGTGACCTGCTACAGCAGGCTCAGGAGATCCGCAAGGTCTTCCGAGCTGCTGCAGCCC AGTTCAGGTGCCCTGGCTGGAGTCAGTCCTCATCGTAGTCAGCAACAACATTGACGA GGAGGCGCTGGCCCGACTGGCCAGGAGGGCAGTGAGGTGAATGTCTATTGGCATTGGC ACCACTGTGGTCACCTGCCCCAACAGCCTTCCCTGGGTGGCGTCTATAAGCTGGTGG CCGTGGGGGGCCAGCCACGAATGAAGCTGACCGAGGACCCCGAGAAGCAGACGTTGCC TGGGAGCAAGGCTGCTTTCCGGCTCCTGGGCTCTGACGGGTCTCCACTCATGGACATG CTGCAGTTAGCAGAAGAGCCAGTGCCACAGGTGGGCAGGAGCTGAGGGTGTGGCCTC CAGGGGGCCAGGAGCCCTGCACCGTGAGGCCAGCCAGGTGGAGCCACTACTGCGGCT CTGCCTCCAGCAGGGACAGCTGTGTGAGCCGCTCCCATCCCTGGCAGAGTCTAGAGCC TTGGCCCAGCTGTCCCTGAGCCGACTCAGCCCTGAGCACAGGCGGCTGCGGAGCCCTG CACAGTACCAGGTGGTGTGTCCGAGAGGCTGCAGGCCCTGGTGAACAGTCTGTGTGC GGGGCAGTCCCCCTCGAG		
	ORF Start: AGA at 1	ORF Stop: at 1180	
	SEQ ID NO: 70	393 aa	MW at 41825.4kD
NOV16c, 174308429 Protein Sequence	RSTNAARVRLIAGPEKRLLEMLRRAQGPDGGLTASTYSYLGFDSSSNVLAGQLRGV PVAGTLAHSFVTSFSGSEVPDPMLAPAAGEGPGVDLAAKAQVWLEQVCAHLGLGVQE PHPGERAAAFVAYALAFPRAFQGLLDITYSVWRSGLPNFLAVALALGELGYRAVGVRDLS GDLLQQAQEIIRKVFRAAAAFQVPWLESVLIVSNNIDEEALARLAQEGSEVNVIGIG TSVVTCPQQPSLGGVYKLVAVGGQPRMKLTEDPEKQTLPGSKAAFRLLGSDGSPMDM LQLAEEPVPQAGQELRVWPPGAQEPCTVRPAQVEPLLRCLCQQQLCEPLPSLAESRA LAQLSLRSLSPHRRRLRSPAQYQVVLSERLQALVNSLCAGQSPLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

<b>Table 16B. Comparison of NOV16a against NOV16b through NOV16c.</b>		
<b>Protein Sequence</b>	<b>NOV16a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV16b	143..531	351/391 (89%)
	2..391	353/391 (89%)
NOV16c	143..531	350/391 (89%)
	2..391	352/391 (89%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

<b>Table 16C. Protein Sequence Properties NOV16a</b>	
<b>PSort analysis:</b>	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.2864 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
<b>SignalP analysis:</b>	No Known Signal Sequence Predicted

- 5 A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 16D.

<b>Table 16D. Geneseq Results for NOV16a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV16a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAG33687	Arabidopsis thaliana protein fragment SEQ ID NO: 40861 - Arabidopsis thaliana, 553 aa. [EP1033405-A2, 06-SEP-2000]	14..531 4..548	231/547 (42%) 330/547 (60%)	e-113
AAG33686	Arabidopsis thaliana protein fragment SEQ ID NO: 40860 - Arabidopsis thaliana, 574 aa. [EP1033405-A2, 06-SEP-2000]	14..531 25..569	231/547 (42%) 330/547 (60%)	e-113
AAG33685	Arabidopsis thaliana protein fragment SEQ ID NO: 40859 - Arabidopsis thaliana, 591 aa. [EP1033405-A2, 06-SEP-2000]	14..531 42..586	231/547 (42%) 330/547 (60%)	e-113

AAY74114	Human prostate tumor EST fragment derived protein #301 - Homo sapiens, 223 aa. [DE19820190-A1, 04-NOV-1999]	334..531 26..223	197/198 (99%) 198/198 (99%)	e-109
AAG29216	Arabidopsis thaliana protein fragment SEQ ID NO: 34723 - Arabidopsis thaliana, 435 aa. [EP1033405-A2, 06-SEP-2000]	14..474 4..432	200/468 (42%) 278/468 (58%)	1e-95

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BRG0	HYPOTHETICAL 58.1 KDA PROTEIN - Homo sapiens (Human), 542 aa (fragment).	1..531 5..542	514/539 (95%) 516/539 (95%)	0.0
Q9VQX4	CG3714 PROTEIN - Drosophila melanogaster (Fruit fly), 541 aa.	14..531 13..536	234/525 (44%) 330/525 (62%)	e-120
O80459	AT2G23420 PROTEIN - Arabidopsis thaliana (Mouse-ear cress), 574 aa.	14..531 25..569	231/547 (42%) 330/547 (60%)	e-112
AAK68525	HYPOTHETICAL 57.8 KDA PROTEIN - Caenorhabditis elegans, 511 aa.	13..445 9..449	198/443 (44%) 290/443 (64%)	e-101
Q95XX1	HYPOTHETICAL 59.9 KDA PROTEIN - Caenorhabditis elegans, 531 aa.	13..445 29..469	198/443 (44%) 290/443 (64%)	e-101

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

5

**EXAMPLE 17.**

The NOV17 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 71	572 bp	
NOV17a, CG59368-01 DNA Sequence	CCGTGGTGCA CGCGCTGCCCCGCATCAACCGCATGGTGCTGTGCTACCTCATCCGCTT CCTGCAGGTCTTCGTGCAGCCGGCCAACGTCGCGGTACCAAGATGGATGTCAGCAAC CTGGCCATGGTGATGGCGCCCAACTGCTTGCGCTGCCAGTCCGACGACCCGCGCGTCA TCTTCGAGAAACCCGCAAGGAGATGTCCTTCCTGCGGGTGCTCATCCAGCACCTGGA CACCAGCTTCATGGAGGGTGTGCTGTAGCGGGGGCGCCCGGGGACAGGAGGGATGTCC TGCCGCCCCCAGCCAGGCCGAACCTCCGCACTCGCTCTCCCGGCAGAGGGGTGAGAAATC GCCCGGCCAGCCCTGGAGCCCCCTCCACTCCCCAGGCCCTGGCCCGGGCGCTCCC CACGCTCTTCTGCCTGGTCTGAGGGTGTAGCCAGGGCACAGCAGCGCGGGGAGGGCGC CTCTGGCCCCCACCTCACGGCCAGTTCCTCGCGGGCACCGCCTCGCCCTCCGCTGGCC GCGGGTCAGCTCCGAGAAAGTGCCTTCTGTAGCTTCATTTATATTAATT		
	ORF Start: ATG at 33	ORF Stop: TAG at 258	
	SEQ ID NO: 72	75 aa	MW at 8638.2kD
NOV17a, CG59368-01 Protein Sequence	MVL CYLIRFLQVFVQPANVAVTKMDVSNLAMVMAPNCLRCQSDDPRVIFENTRKEMSF LRVLIQHLDTSFMEGV L		

Further analysis of the NOV17a protein yielded the following properties shown in Table 17B.

Table 17B. Protein Sequence Properties NOV17a	
PSort analysis:	0.8134 probability located in mitochondrial intermembrane space; 0.5255 probability located in mitochondrial matrix space; 0.2672 probability located in lysosome (lumen); 0.2537 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 20 and 21

- A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE03048	Human preoptic regulatory factor-2 (hPORF-2) protein #1 - Homo sapiens, 75 aa. [WO200142464-A2, 14-JUN-2001]	1..75 1..75	75/75 (100%) 75/75 (100%)	1e-37

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17D.

Table 17D. Public BLASTP Results for NOV17a
---

Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9C0H5	KIAA1688 PROTEIN - Homo sapiens (Human), 1094 aa (fragment).	1..75 1020..1094	75/75 (100%) 75/75 (100%)	4e-37
P18890	Putative preoptic regulatory factor-2 precursor (PORF-2) - Rattus norvegicus (Rat), 75 aa.	1..75 1..75	74/75 (98%) 75/75 (99%)	5e-37
Q9VDE9	CG3421 PROTEIN - Drosophila melanogaster (Fruit fly), 1309 aa.	1..75 1235..1309	48/75 (64%) 58/75 (77%)	2e-21

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17E.

Table 17E. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### EXAMPLE 18.

The NOV18 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
	SEQ ID NO: 73	1452 bp
NOV18a, CG58628-01 DNA Sequence	ATCCTGCCCCGCAGGGTGACCCCTGTTTGACGACGATGTCTGAAGAAGAGGCGGCTCA GATCCCCAGATCCAGTGTGTGGGAGCAGGACCAGCAGAACGTGGTGCAGCGTGTGGTG GCTCTGCCCCCTGGTCAGGGCCACGTGCACCGCGGTCTGCGATGTTTACAGTGCAGCCA AGGACAGGCACCCGCTGCTGGGCTCCGCCTGCCGCTGGCTGAGAACTGCGTGTGCGG CCTGACCACCCGTGCCCTGGACCACGCCAGCCGCTGCTCGAGCACCTGCAGCCCCAG GTGGCCACTATGAACAGCCTCGCCTGCAGGGGCTGGACAAGCTGGAAGAGAAGCTTC CCTTTCTCCAGCAACCTTCGGAGACGCTAGTGACCTCAGCCAAGGACGTGGTGGCCAG CAGTGTACGGGTGTGGTGACCTGGCCCGAGGGGCCGCGCTGGAGCGTGGAGCTG AAGCGTCCGTGAGCCATGCTGTGGATGTTGTACTGGAATAATCAGAGGAGCTGGTGG ATCACTTCCTGCCCATGACGAGGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGG CCCTGAAGTGGGTTTCGGTGGAGGATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTC GGCTCCCTGTACAGCAGGATCCGCCACCTGGCCTACGAGCACTCTGTGGGGAACTGA GGCAGAGCAAACACCGTGCCAGGACACCCTGGCCAGCTGCAGGAGACGCTGGAGCT GATAGACCACATGCAGTGTGGGTGACCCCAACCGCCCGGCTGCCCTGGGAAGGTG CACGAGCTGTGGGGGAATGGGGCCAGCGCCCTCCGGAGAGCCGCCCGGAGCCAGG TGGAGCTGGAGACGCTGGTGTGTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGGT AGAGGCTCTGGAGTCCAGCGTGCGGGGCCTGCCCGCCGGCGCCAGGAGAAGGTGGCT GAGGTGCGGCGCAGTGTGGATGCCCTGCAGACCGCCTTCGCTGATGCCCGCTGCTTCA GGGACGTGCCAGCGCCGCGCTGGCCGAGGGCCGGGTCGCGTGGCCACCGCAGCGC CTGCGTGGACGAGCTGCTGGAGCTGGTGGTGCAGGCCGTGCCGCTGCCCTGGCTGGTG GGACCCTTCGCGCCCATCCTTGTGGAGCGACCCGAGCCCTGCCGACCTGGCGGACC	



	TGGTGGACGAGGTTCATCGGGGGCCCTGACCCCCGCTGGGCGCACCTGGACTGGCCGGC CCAGCAGAGAGCCTGGGGAGCAGACAGGGACGGGAGTGGGAATGGGGATGGGGAC AGGATGGGTGTTGCCGGGGACATCTGCGAGCAGGAACCCGAGACCCCGAGCTGCCCGG TCAAGCACACCCTGATGCCCGAGCTGGACTTCTGACCCATGGGCCAGTGGAGGCGGGG AG		
	ORF Start: ATG at 36	ORF Stop: TGA at 1425	
	SEQ ID NO: 74	463 aa	MW at 50804.9kD
NOV18a, CG58628-01 Protein Sequence	MSEEEAAQIPRSSVWEQDQQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACR LAENCVCGLTTRALDHAQPLLEHLQPQVATMNSLACRGLDKLEEKLPFLQQPSETVVT SAKDVVASSVTGVVDLARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEEELA ALAAEAEGPEVGSVEDQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDTLA QLQETLELIDHMOCGVTPAPACPGKVHELWGEWQRPPESSRRRSQVELETLVLSRSL TQELQGTVEALESSVRGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVPAALAEGR GRVAHAHACVDELLELVQAVPLPWLVGPFAPILVERPEPLDLADLVDEVIGGPDP WAHLDPAAQQRAWEAEHRDGSNGDGDGMVAGDICEQEPETPSPVKHTLMPELDF		
	SEQ ID NO: 75	978 bp	
NOV18b, 174228350 DNA Sequence	AGATCTGACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCCGGTTCAGGGCCA CGTGCAACCGCGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGG CTCCGCTGCGCGCTGGCTGAGAAGTGCAGTGTGCGGCTGACACCCGCTGCCCTGGAC CACGCCAGCCGCTGCTCGAGCACCTGCAGCCCAGCTGGCCACTATGAACAGCCTCG CCTGCAGGGGCTCGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGA GACGGTGGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACAGGGTGTGGTGGAC CTGGCCCGGAGGGGCCGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTG TGGATGTTGTACTGGAATAATCAGAGGAGCTGGTGGATCACTTCTGCCCATGACGGA GGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGTTTCGGTGGAG GATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTGACACGGATCC GCCACCTGGCCTACGAGCACTCTGTGGGGAAGTGAAGGACAGCAAAACCCGTGCCCA GGACACCCTGGCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGG GTGACCCCCACCGCCCCGCGCCCTGGGAAGGTGCACGAGCTGTGGGGGAATGGG GCCAGCGCCCTCCGAGAGCCGCCCGGAGCCAGGCAGAGCTGGAGACGCTGGTGTCT GTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTG CGGGGCTGCCCCGCGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATG CCCTGCAGACCGCTTCGCTGATGCCCGCTGCTTCAGGACGCTGGTTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 76	326 aa	MW at 35954.4kD
NOV18b, 174228350 Protein Sequence	RSDQQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACRLAENCVCGLTTRALD HAQPLLEHLQPQLATMNSLACRGLDKLEEKLPFLQQPSETVVTSAKDVVASSVTGVVD LARRGRRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTEEEELAALAAEAEGPEVGSVE DQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDTLAQLQETLELIDHMOCG VTPAPARPGKVHELWGEWQRPPESSRRRSQAELETLVLSRSLTQELQGTVEALESSV RGLPAGAQEKVAEVRRSVDALQTAFADARCFRDVVD		
	SEQ ID NO: 77	978 bp	
NOV18c, 174228354 DNA Sequence	AGATCTGACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCCGGTTCAGGGCCA CGTGCAACCGCGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGG CTCCGCTGCGCGCTGGCTGAGAAGTGCAGTGTGCGGCTGACACCCGCTGCCCTGGAC CACGCCAGCCGCTGCTCGAGCACCTGCAGCCCAGCTGGCCACTATGAACAGCCTCG CCTGCAGGGGCTCGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGA GACGGTGGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACAGGGTGTGGTGGAC CTGGCCCGGAGGGGCCGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTG TGGATGTTGTACTGGAATAATCAGAGGAGCTGGTGGATCACTTCTGCCCATGACGGA GGAAGAGCTCGCGGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGTTTCGGTGGAG GATCAGAGGAGACAGCAGGGCTACTTTGTGCGCCTCGGCTCCCTGTGACACGGATCC GCCACCTGGCCTACGAGCACTCTGTGGGGAAGTGAAGGACAGCAAAACCCGTGCCCA GGACACCCTGGCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGG GTGACCCCCACCGCCCCGCGCCCTGGGAAGGTGCACGAGCTGTGGGGGAATGGG		

	GCCAGCGCCCTCCGGAGAGCCGCCGCCGAGCCAGGCAGAGCTGGAGACGCTGGTGCTGTCCCGCAGCCTGACCCAGGAGCTGCAGGGCACGCTAGAGGCTCTGGAGTCCAGCGTGTGGGGCTGCCCGCCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATGCCCTGCAGACCGCCTTCGTGATGCCCGCTGCTTCAGGGACGTGGTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 78	326 aa	MW at 35984.4kD
NOV18c, 174228354 Protein Sequence	RSDQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACRLAENCVCGLTTRALDHAQPLLEHLQPQLATMNSLACRGLDKLEKLPFLQQPSETVVTSAKDVVASSVTGVVDLARRGRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTTEELAALAAEAGPEVGSVEDQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDTLAQLQETLELIDHMCCGVTPTAPARPGKVHELWGEWQRPPESSRRRSQAELETLVLSRSLTQELQGTVEALESSVWGLPAGAQEKVAEVRSSVDALQTAFADARCFRDVVD		
	SEQ ID NO: 79	1401 bp	
NOV18d, 188822733 DNA Sequence	AGATCTATGTCTGAAGAAGAGGCGGCTCAGATCCCCAGATCCAGTGTGTGGGAGCAGGACCAGCAGAACGTGGTGCAGCGTGTGGTGGCTCTGCCCTGGTCAGGGCCACGTGCACCGCGTCTGCGATGTTTACAGTGCAGCCAAGGACAGGCACCCGCTGCTGGGCTCCGCCGTGCCGCTGGCTGAGAACTGCGTGTGCGGCCTGACCACCCGTGCCCTGGACCACGCCCAGCCGCTGCTCGAGCACCTGCAGCCCCAGCTGGCCACTATGAACAGCCTCGCCTGCAGGGCCTGGACAAGCTGGAAGAGAAGCTTCCCTTTCTCCAGCAACCTTCGGAGACGGTGTGTGACCTCAGCCAAGGACGTGGTGGCCAGCAGTGTACGGGTGTGGTGGACCTGGCCCGGAGGGCCGGCGCTGGAGCGTGGAGCTGAAGCGCTCCGTGAGCCATGCTGTGGATGTGTACTGGAATAATCAGAGGAGCTGGTGGATCACTTCCTGCCCATGACGGAGGAAGAGCTCCTCGCGCACTGGCGGCTGAGGCTGAAGGCCCTGAAGTGGGTTCGGTGGAGGATCAGAGGAGACAGGGCTACTTTGTGCGCCTCGGCTCCCTGTCAGCACGGATCCGCCACCTGGCCTACGAGCACTCTGTGGGAACTGAGGCAGAGCAAAACACCGTGCCAGGACACCCCTGGCCAGCTGCAGGAGACGCTGGAGCTGATAGACCACATGCAGTGTGGGTGACCCCCACCGCCCCGGCCCGCCTGGGAAGGTGCACGAGCTGTGGGGGAATGGGGCCAGCGCCCTCCGGAGAGCCGCCCGGAGCCAGGCAGAGCTGGAGACGCTGGTGCTGTCCCCGAGCCTGACCCAGGAGCTGCAGGGCACGGTAGAGGCTCTGGAGTCCAGCGTGTGGGGCCGTGCCCGCGCGCCAGGAGAAGGTGGCTGAGGTGCGGCGCAGTGTGGATGCCCTGCAGACCGCCTTCGTGATGCCCGCTGCTTCAGGGACGTGCCAGCGCGCGCTGGCCGAGGGCCGGGTTCGCTGGCCACGCGCACGCCTGCGTGGACGAGCTGCTGGAGCTGGTGGTGCAGGCGTGCCTGCTGCCCTGGCTGGTGGGACCTTCGCGCCCATCCTGTGGAGCGACCCGAGCCCCTGCCGACCTGGCGGACCTGGTGGACGAGGTCATCGGGGGCCCTGACCCCGCTGGGCGCACCTGGACTGGCCGGCCAGCAGAGAGCCTGGGAGGCAGAGCACAGGGACGGGATGGGAATGGGATGGGGACAGGATGGGTGTTGCCGGGGACATCTGCCAGGCAGGAACCCGAGACCCCGAGCTGCCCGGTCAAGCACACCCTGATGCCCGAGCTGGACCTTCGTCGAC		
	ORF Start: AGA at 1	ORF Stop:	
	SEQ ID NO: 80	467 aa	MW at 51331.4kD
NOV18d, 188822733 Protein Sequence	RSMSEEEAAQIPRSSVWEQDQNVVQRVVALPLVRATCTAVCDVYSAAKDRHPLLGSACRLAENCVCGLTTRALDHAQPLLEHLQPQLATMNSLACRGLDKLEKLPFLQQPSETVVTSAKDVVASSVTGVVDLARRGRWSVELKRSVSHAVDVVLEKSEELVDHFLPMTTEELAALAAEAGPEVGSVEDQRRQQGYFVRLGSLSARIRHLAYEHSVGLRQSKHRAQDTLAQLQETLELIDHMCCGVTPTAPARPGKVHELWGEWQRPPESSRRRSQAELETLVLSRSLTQELQGTVEALESSVWGLPAGAQEKVAEVRSSVDALQTAFADARCFRDVPAAALAEGRGRVAHAHACVDELLELVQAVPLPWLVPFPAPILVERPEPLPDADLVDEVIGGPDPRWAHLDPAAQRAWAEHRDGSNGDGRMGVAGDICEQEPETPSCPVKHTLMPELDFVD		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

**Table 18B. Comparison of NOV18a against NOV18b through NOV18d.**

Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	18..339 3..324	307/322 (95%) 308/322 (95%)
NOV18c	18..339 3..324	306/322 (95%) 307/322 (95%)
NOV18d	1..463 3..465	447/463 (96%) 448/463 (96%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY67240	Human adipophilin-like protein (HALP) amino acid sequence - Homo sapiens, 434 aa. [US5989820-A, 23-NOV-1999]	19..385 22..428	163/407 (40%) 234/407 (57%)	1e-77
AAW59883	Amino acid sequence of the cDNA clone ADF (HFKFY79) - Homo sapiens, 452 aa. [WO9831800-A2, 23-JUL-1998]	19..388 22..431	149/411 (36%) 223/411 (54%)	4e-64
AAM25962	Human protein sequence SEQ ID NO:1477 - Homo sapiens, 139 aa. [WO200153455-A2, 26-JUL-2001]	1..117 23..139	116/117 (99%) 117/117 (99%)	4e-62
AAY99534	Human adipocyte-specific differentiation-related protein ADRP - Homo sapiens, 437 aa. [WO200031532-A1, 02-JUN-2000]	12..384 2..411	140/416 (33%) 222/416 (52%)	1e-59

AAW53264	Human adipocyte-specific differentiation-related protein - Homo sapiens, 437 aa. [US5739009-A, 14-APR-1998]	12..384 2..411	140/416 (33%) 222/416 (52%)	1e-59
----------	---	-------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a				
Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D6M0	2310076L09RIK PROTEIN - Mus musculus (Mouse), 448 aa.	1..463 1..448	329/463 (71%) 368/463 (79%)	0.0
Q9BS03	CARGO SELECTION PROTEIN (MANNOSE 6 PHOSPHATE RECEPTOR BINDING PROTEIN) - Homo sapiens (Human), 434 aa.	19..385 22..428	163/407 (40%) 234/407 (57%)	4e-77
O60664	Cargo selection protein TIP47 (47 kDa mannose 6-phosphate receptor- binding protein) (47 kDa MPR-binding protein) (Placental protein 17) - Homo sapiens (Human), 434 aa.	19..385 22..428	163/407 (40%) 234/407 (57%)	6e-77
Q9DBG5	1300012C15RIK PROTEIN (RIKEN CDNA 1300012C15 GENE) - Mus musculus (Mouse), 437 aa.	19..385 22..432	160/411 (38%) 232/411 (55%)	4e-73
Q9CZK1	1300012C15RIK PROTEIN - Mus musculus (Mouse), 437 aa.	19..385 22..432	160/411 (38%) 232/411 (55%)	6e-73

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Man-6-P_recep: domain 1 of 1	156..168	9/13 (69%) 9/13 (69%)	0.7
perilipin: domain 1 of 1	10..369	139/411 (34%) 240/411 (58%)	1.4e-76

**EXAMPLE 19.**

The NOV19 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 81	774 bp	
NOV19a, CG59342-01 DNA Sequence	GTAGAGTTTTTCAGGTTGCTCCTGGAAACCATGCCGAAAGTAGTGTCTCGGTCAGTAG TCTGCTCTGACACTCGGGACCGGGAGGAATATGACGACGGCGAGAAGCCCTCCATGT GTACTACTGTTTGTGCGGCCAGGTGGTCTCTAGTGTCTGGACTGTCTAGTTAGAGAAATTG CCCATGAGGCCCCCGGGACCGGTCCCGTGTGATTGATGCTGCCAAACATGCCCATAGT TTTGTAAACACAGAAGACGAAGAGACTATGTATCTGCGGAGACCTGAAGGCATTGAAT ACAGTACAGAAAGAAATGTGCAAAAGTGTGGACTGCTGCTCTTCTACCAATCCCAGCCG AAGAATGCTCCCGTTACCTTCATTGTGGATGGAGCAGTCGTCAAGTTTGGCCAGGGCT TTGGGAAAACGAACATATATACTCAGAAACAAGAGCCTCCTAAGAAGGTGATGATGAC CAAACGGACCAAAGACATGGGCAAGTTCAGTTCGTCACTGTGTCTACCATTGATGAA GAGGAAGAGGAGATTGAGGCTAGGGAAGTTTGCTGACTCGTATGCACAGAATGCCAAAC TGATTGAAAAACAGCTGGAGCGCAAAGGCATGAGCAAGAGGCCACTGCAAGAGCTGGC TGAATGGGAACCCAGGAAAAGAGGACATATGACACAGGTTCTCCCTCTGCAAAAAAG TGGCAGATGCGTGGCTCAGGGGCCTTCCACTGTCCAGGTCTCCTCAGATGGCCCTGG GAATGAGCGGCCACCATTAA		
	ORF Start: ATG at 31	ORF Stop: TAA at 772	
	SEQ ID NO: 82	247 aa	MW at 28211.1kD
NOV19a, CG59342-01 Protein Sequence	MPKVVSRSVVCSDTRDREYDDGEKPLHVYYCLCGQVVLVLDQLEKLEPMRPRDRSRV IDAAKHAKFCNTEDEETMYLRRPEGIELQYRKKCAKCGLLFYQSQPKNAPVTFIVL GAVVKFGQGFGKTNITYTQKQEPKKVMMTKRRTKDMGKFSSVTVSTIDEEEEIEAREV ADSYAQNAKVIKQLERKGMKRPLQLAWEWPEKRTYDTGSPSAKKWQMRGSGAFH CPGPPQMALGMSGHH		

Further analysis of the NOV19a protein yielded the following properties shown in

5 **Table 19B.**

Table 19B. Protein Sequence Properties NOV19a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3600 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 19C.

Table 19C. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAG01784	Human secreted protein, SEQ ID NO: 5865 - Homo sapiens, 87 aa. [EP1033401-A2, 06-SEP-2000]	1..86 1..86	85/86 (98%) 86/86 (99%)	3e-46
AAM41425	Human polypeptide SEQ ID NO 6356 - Homo sapiens, 92 aa. [WO200153312-A1, 26-JUL-2001]	139..215 9..85	65/77 (84%) 69/77 (89%)	4e-28
AAM39639	Human polypeptide SEQ ID NO 2784 - Homo sapiens, 80 aa. [WO200153312-A1, 26-JUL-2001]	143..215 1..73	63/73 (86%) 66/73 (90%)	6e-27
AAG60283	Arabidopsis thaliana protein fragment SEQ ID NO: 78065 - Arabidopsis thaliana, 236 aa. [EP1033405-A2, 06-SEP-2000]	1..119 1..122	44/127 (34%) 64/127 (49%)	2e-09
AAG59843	Arabidopsis thaliana protein fragment SEQ ID NO: 77448 - Arabidopsis thaliana, 230 aa. [EP1033405-A2, 06-SEP-2000]	1..119 1..116	42/123 (34%) 63/123 (51%)	8e-09

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19D.

Table 19D. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H5V9	CDNA: FLJ22965 FIS, CLONE KAT10418 - Homo sapiens (Human), 222 aa.	1..215 1..215	202/215 (93%) 206/215 (94%)	e-114
AAH21479	HYPOTHETICAL 25.6 KDA PROTEIN - Mus musculus (Mouse), 222 aa.	1..215 1..215	201/215 (93%) 205/215 (94%)	e-113
Q9CWC1	C330007P06RIK PROTEIN - Mus musculus (Mouse), 250 aa.	1..202 1..202	197/202 (97%) 198/202 (97%)	e-111
Q9V412	BG:DS00941.3 PROTEIN - Drosophila melanogaster (Fruit fly), 247 aa.	1..193 1..218	106/220 (48%) 145/220 (65%)	2e-50
Q95Q06	Y66D12A.8 PROTEIN - Caenorhabditis elegans, 244 aa.	13..194 29..207	79/187 (42%) 114/187 (60%)	1e-30

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19E.

Table 19E. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

### EXAMPLE 20.

The NOV20 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis			
	SEQ ID NO: 83	324 bp	
NOV20a, CG59486-01 DNA Sequence	ATTTTGTGTTATTGTTGTAGATATGTGGTTTCCCCATGTTGCCAGCTGGCCTC AACTCCTGGCCTCAAGATCCACCCGCCTCGACCTCCCAAAGGCCAGCCCCCTCTTT CCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTGTTTTAAAAAAAAAA AAAGGCCAGGCGCAGTGGCTCATGTCTGTAATCCCAGCACTCTGGGAGGCCAAGGCA GCAGATCACAAAGTCAGGAGATCAAGACCATCTGGCTAACACAGTGAAACCCATCT CTACTAAAAATACAAAAAAATTAGCCAGGCG		
	ORF Start: ATG at 40	ORF Stop: TAA at 295	
	SEQ ID NO: 84	85 aa	MW at 9476.2kD
NOV20a, CG59486-01 Protein Sequence	MLPAGLELLASRSTRLDLPKAQPLSFLPSFLPSFLPSFLVFKKKKKGQAQWLMSVIP LWEAKAGRSQGQEIKTILANTVKPHLY		

**Further analysis of the NOV20a protein yielded the following properties shown in**

5 Table 20B.

Table 20B. Protein Sequence Properties NOV20a	
PSort analysis:	0.6238 probability located in microbody (peroxisome); 0.6000 probability located in nucleus; 0.3600 probability located in mitochondrial matrix space; 0.1830 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 20C.

Table 20C. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAB95050	Human protein sequence SEQ ID NO:16847 - Homo sapiens, 112 aa. [EP1074617-A2, 07-FEB-2001]	35..85 62..112	38/51 (74%) 42/51 (81%)	4e-15
ABB11422	Human Zn finger protein homologue, SEQ ID NO:1792 - Homo sapiens, 670 aa. [WO200157188-A2, 09-AUG-2001]	47..85 632..670	32/39 (82%) 35/39 (89%)	1e-12
AAM85296	Human immune/haematopoietic antigen SEQ ID NO:12889 - Homo sapiens, 81 aa. [WO200157182-A2, 09-AUG-2001]	37..84 19..67	35/49 (71%) 41/49 (83%)	1e-12
AAM94124	Human reproductive system related antigen SEQ ID NO: 2782 - Homo sapiens, 107 aa. [WO200155320-A2, 02-AUG-2001]	41..85 63..107	33/45 (73%) 38/45 (84%)	3e-12
AAM91494	Human immune/haematopoietic antigen SEQ ID NO:19087 - Homo sapiens, 58 aa. [WO200157182-A2, 09-AUG-2001]	49..85 22..58	32/37 (86%) 34/37 (91%)	3e-12

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20D.

Table 20D. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UI59	PRO0478 - Homo sapiens (Human), 87 aa.	8..85 7..87	50/82 (60%) 59/82 (70%)	4e-18
P39189	Alu subfamily SB sequence contamination warning entry - Homo sapiens (Human), 587 aa.	47..85 1..39	30/39 (76%) 35/39 (88%)	1e-10
P39192	Alu subfamily SC sequence contamination warning entry - Homo sapiens (Human), 585 aa.	47..85 1..39	29/39 (74%) 34/39 (86%)	5e-10
P39191	Alu subfamily SB2 sequence contamination warning entry - Homo sapiens (Human), 603 aa.	47..85 1..39	29/39 (74%) 33/39 (84%)	9e-10
P39190	Alu subfamily SB1 sequence contamination warning entry - Homo sapiens (Human), 587 aa.	47..85 1..39	28/39 (71%) 33/39 (83%)	3e-09



Pfam analysis predicts that the NOV20a protein contains the domains shown in the Table 20E.

Table 20E. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### EXAMPLE 21.

The NOV21 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 85	1572 bp	
NOV21a, CG59446-01 DNA Sequence	GGTGTGCAGGATATAAGGTTGGACTTCCAGACCCACTGCCCGGGAGAGGAGAGGAGCG GGCCGAGGACTCCAGCGTGCCCAGGTCTGGCATCCTGCACCTTGCTGCCCTCTGACACG TGGGAAGATGGCCGGCCCCGTGGACCTTCACCCTTCTCTGTGGTTTGCTGGCAGCCACG TTGATCCAAGCCACCCTCAGTCCCCTGTCAGTTCTCATCCTCGGCCCAAAAGTCATCA AAGAAAAGCTGACACAGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCT GCCGCTGCTCAGTGCCATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGC CTGGTGAACACCGTCTCTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCG TCCAGCTGCAGGTGAAGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCT GGACATGGTGGCTGGATTCAACACGCCCTGGTCAAGACCATCGTGGAGTTCCACATG ACGACTGAGGCCCAAGCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCC TGGTCCTCAGTGACTGTGCCACCAGCCATGGGAGCCTGCGCATCCAACCTGCTGCATAA GCTCTCCTTCTGGTGAACGCCTTAGCTAAGCAGGTATGAACCTCCTAGTGCCATCG CTGCCAATCTAGTGAAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCA TGTATGCAGACCTCCTGCAGCTGGTGAAGGTGCCATTTCCTCAGCATTGACCGCT GGAGTTTGACCTTCTGTATCTTGCCATCAAGGGTGACACCATTCAGTCTACTTGGGG GCCAAGTTGTTGACTCACAGGGAAGGTGACCAAGTGGTTCAATAACTCTGCAGCT CCCTGACAATGCCACCCCTGGACAACATCCCCTTCAGCCTCATCGTGAGTCAACCGTT CAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGTGGCTGCTGTGCTCTCTCC GAAGAATTCATGGTCTGTGGACTCTGTGCTTCTGAGAGTGCCCATCGGCTGAAGT CAAGCATCGGGCTGATCAATGAAAAGGAAGCCAGCTCGGAAGCTCAGTTTACACCA AGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCTGATG AACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAGATCA TCCACTCCATCCTGTGCGGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAGTGT ATTGGTGAAGGCCTTGGGATTCGAGGCAGCTGAGTCTCTCACTGACCAAGGATGCCCT GTGCTTACTCCAGCCTCCTTGTGGAACCCAGCTCTCCTGTCTCCAGTGAAGACTG GATGGCAGCCATCAGGGAAGGCTGGGTCCCAGTTGGGAGTATGGGTGTGAGCTCTATA GACCATCCCTCTCTGCAATCAATAAACACTTGCTGTGAAAAAAAAAAAAAAAAATAAAA AAAAAA		
	ORF Start: ATG at 124	ORF Stop: TGA at 1441	
	SEQ ID NO: 86	439 aa	MW at 47572.2kD
NOV21a, CG59446-01 Protein Sequence	MAGPWTFTLLCGLLAATLIQATLSPTAVLILGPKVIEKLTQELKDHNATSILQQLP LSAMREKPAGGIPVLGSLVNTVLKHIWLKVITANILQLQVKPSANDQELLVKIPLDN VAGFNTPLVKTI VEFHMTTEAQATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLS FLVNALAKQVMNLLVPSLPNLVKNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLE DLLYPAIKGDTIQLYLGAKLLDSQGVTKWFNNSAASLTMPITLDNIPFSLIVSHPPSL IVSQDVVKA AVAVLSPEEFMVLLDSVLPESAHLKSSIGLINEKEASSEAQFYTKG QLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITEIIHSILLPNQNGKLRSQVPVSL		

	KALGFEEAESSLTkdALVLTpASLWKpSSpVSQ		
	SEQ ID NO: 87	1392 bp	
NOV21b, 174308261 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCTCGGCCCAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAACACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAATGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCGCTCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGACCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTCAACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTGAGGCGAGCTGAGTCCCTCACTGACCAAGGATGC CCTTGCTGTACTCCAGCCTCCTTGTGGAAACCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop: LV at 1393	
	SEQ ID NO: 88	464 aa	MW at 50459.5kD
NOV21b, 174308261 Protein Sequence	KLPTAVLILGPKVIKEKLQELKDHNATSILQQLPLLSAMREKPAAGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLDFDLYPAIKGDTIQLYLGAKLLD SQGKVTWKFNNASASLTMTPLDNIPIFSLIVSQDVVKAAVAVALSPFEFVLLDSVLPE SAHRLKSSIGLINEKAADKLQPTQIVKILTQDTPFEFFIDQGHAKVAQLIVLEVPFSE ALRPLFTLGIIEASDAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQKLRSGVPVSLVKALGFEEAESSLTkdALVLTpASLWKpSSpVSQLE		
	SEQ ID NO: 89	1392 bp	
NOV21c, 174308266 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCTCGGCCCAAAGTCATCAAAGAAAAGCCGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCCTGGTCCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAATGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTCAACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT		

	GATGAACTCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACAGAAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 90	464 aa	MW at 50433.4kD
NOV21c, 174308266 Protein Sequence	KLPTAVLILGPKVIKEKPTQELKDNATSILOQLPLLSAMREKPAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDTIQLYLGAKLLD SQGKVTWKFNNASASLTMPITLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKL GSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFPSS ALRPLFTLGIEASSEAQFYTKGDQLILNLNINISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQL		
	SEQ ID NO: 91	1392 bp	
NOV21d, 174308278 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTCT CTGAAGCACATCATCTGGCTGAAGGTATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCACATGGACACCAGTGCAAGTGGCCCCACCCGCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCTTGGT GAACGCCTTAGCTAAGCAGGTATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTCAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCCTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGTGGAAGTGTTCCTCCAGTGAA GCCCTCCGCCCTTTGTTCAACCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCTTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 92	464 aa	MW at 50430.4kD
NOV21d, 174308278 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDNATSILOQLPLLSAMREKPAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIHMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDTIQLYLGAKLLD SQGKVTWKFNNASASLTMPITLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKL GSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFPSS ALRPLFTLGIEASSEAQFYTKGDQLILNLNINISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPSSPVSQL		
	SEQ ID NO: 93	1392 bp	
NOV21e, 174308283 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTCT CTGAAGCACATCATCTGGCTGAAGGTATCACAGCTAACATCCTCCAGTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTTAAGATCCCCCTGGACATGGTGGCTGG		

	ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAAGTGCAAGTGGCCCCACCCGCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAACGCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCCTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATCAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCCCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG CGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCTCTACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 94	464 aa	MW at 50431.4kD
NOV21e, 174308283 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDNATSIQLQLPLLSAMREKPAAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIVEFHMTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKL SFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDITQIYLGAKLDD SQGKVTWKFNNASAASLTMPDLNIPFSLIVSQDVVKAAVAVALSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPFFIDQGHAKVAQLIVLEVFPSSSE ALRPLFTLGI EASSEAQFYTKGDQLILNLNISPDRIQLMNSIGWFPDVLKNIITE I IHSILLENQNGKLRSGVPASLVKALGFEEAESSLT KDALVLT PASLWKPPSPVSQLE		
	SEQ ID NO: 95	1392 bp	
NOV21f, 174308287 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAGGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTCT CTGAAGCACATCATCTGGCTGAAGGTTCATCACAGTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAAGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAAGTGCAAGTGGCCCCACCCGCTGGTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAACCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCCTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCGTTTTCAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTATAGA CCAGGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCTCTACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 96	464 aa	MW at 50435.4kD

NOV21f, 174308287 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDNHATSILQQLPILLSAMREKPAGGIPVLGSLVNTV LKHHIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYPAIKGDTVQLYLGAKLLD SQGKVTWKFNNASAASLTMPITLDNIPFSLIVSQDVVKA AVAAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPPEFFIDQGHAKVAQLIVLEVFPSS ALRPLFTLGI EASSEAQFYTKGDQLILNLNINISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILPNQNGKLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPPSPVSQL E		
	SEQ ID NO: 97	1392 bp	
NOV21g, 174308293 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGAAGTGGCCCCACCCGCTGGTCTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTACGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAGATTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTCCTCCAGTGAA GCCCTCCGCCCTTGTTCACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTACA CCAAAGGTGACCAACTTATACTCAACTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAGATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 98	464 aa	MW at 50477.5kD
NOV21g, 174308293 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDNHATSILQQLPILLSAMREKPAGGIPVLGSLVNTV LKHHIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTI VEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYPAIKGDTVQLYLGAKLLD SQGKVTWKFNNASAASLTMPITLDNIPFSLIVSQDVVKA AVAAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPPEFFIDQGHAKVAQLIVLEVFPSS ALRPLFTLGI EASSEAQFYTKGDQLILNLNINISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILPNQNGRLRSGVPVSLVKALGFEEAESSLT KDALVLT PASLWKPPSPVSQL E		
	SEQ ID NO: 99	1392 bp	
NOV21h, 174308301 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACGTCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGAAGTGGCCCCACCCGCTGGTCTCCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA		

	CCCTGGACAACATCCCCTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACGCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCTCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 100	464 aa	MW at 50418.5kD
NOV21h, 174308301 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGIPVLGSLVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDTIQLYLGAKLLD SQGKVTWKFNNSAASLTMTPLDNIPFSLIVSQDVVKAAVAALVSPPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVPSSSE ALRPLFTLIGIEASSEAQFYTKGDQLILNLNLISSDRIQLMNAIGWFPQDVLKNIITE IIHSILLPNQNGKLRSVPSLVKALGFEEAESSLTkdALVLTpASLWKPPSPVSQLE		
	SEQ ID NO: 101	1392 bp	
NOV21i, 174308311 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCCGTGAACACCGTC CTGAAGCACGTCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGCGCCCCACCCGCTGGTCTCAGTGA GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTGATGAACCTCCTAGTGCCATCCCTGCCAATCTAGTG AAAAACCAGCTGTGTCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCTCAGCATTGGCCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTACGCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTTGGACTCTGTGCTTCTGAG AGTGCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTGAGGCAGCTGAGTCTCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 102	464 aa	MW at 50360.4kD
NOV21i, 174308311 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGIPVLGSPVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIGRLEFDLLYPAIKGDTIQLYLGAKLLD SQGKVTWKFNNSAASLTMTPLDNIPFSLIVSQDVVKAAVAALVSPPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVPSSSE ALRPLFTLIGIEASSEAQFYTKGDQLILNLNLISSDRIQLMNSGIGWFPQDVLKNIITE IIHSILLPNQNGKLRSVPSLVKALGFEEAESSLTkdALVLTpASLWKPPSPVSQLE		

	SEQ ID NO: 103	1392 bp	
NOV21j, 174308315 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTGGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGAAGTGGCCCCACCCGCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTGATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTCAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTATCATGGTCTGTGGACTCTGTGCTTCTGAG AGTGGCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAGGGTGGCCCCAAGTATCGTGCTGGAAGTGTCTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG ATCATCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGATGAGTCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 104	464 aa	MW at 50461.4kD
NOV21j, 174308315 Protein Sequence	KLPTAVLILGPKVIKEKLTQELKDHNATSILQQLPLLSAMREKPAAGGIPVLGSLVNTV LKHI IWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKITVEFHMTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLSFLVNALAKQVNNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLYPAIKGDITQLYLGAKLDD SQGKVTWKFNNASLTMTPLDNIPFSLIVSQDVVKAAVAALSPPEFVMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPFEFFIDQGHARVAQLIVLEVPSSSE ALRPLFTLGIEASSEAFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSVPSLVKALGFADSSSLTKDALVLTPLASLWKPPSPVSQLE		
	SEQ ID NO: 105	1392 bp	
NOV21k, 174308321 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCATCACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTGGTCAAGACCATCGTGGAGTTCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGAAGTGGCCCCACCCGCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTGATGAACCTCCTAGTGCCATCCCTGCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTCAGCTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTGTGGACTCTGTGCTTCTGAG AGTGGCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCCAAGTATCGTGCTGGAAGTGTTCCTCCAGTGTA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAAACATCATCACTGAG		

	ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 106	464 aa	MW at 50419.5kD
NOV21k, 174308321 Protein Sequence	KLPTAVLILGPKVIKEKLTOELKDNATSILOQLPLLSAMREKPAGGIPVLGSLVNTV LKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKL SFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADLLQLVKVPI SLSIDRLEFDLLYP AIKGDITQLYLGAKLLD SQGKVTWKFNNASAASLTMTPLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPEFFIDQGHAKVAQLIVLEVFPSSV ALRPLFTLGI EASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSVGVPSLVKALGFEESSSLTKDALVLT PASLWKPSPPVSQLE		
	SEQ ID NO: 107	1392 bp	
NOV21l, 174308327 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACATCATCTGGCTGAAGGTCAACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTCGTCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA GCCACCATCCGCATGGACACCAGTGCAAGTGGCCCCACCCGCTGGTCTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAACCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTCAAGACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAAGCTGTGTCCCGTGATCGAGGCTTCCTTCAATGGCATGTATGCAGACCCCC TGCAGCTGGTGAAGGTGCCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GTATCCTGCCATCAAGGGTGACACCATTCAGCTCTACCTGGGGGCCAAGTTGTGGAC TCACAGGGAAAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACATGCCCCA CCCTGGACAACATCCCCTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTATGGTCTCTGTTGGA CTCTGTGCTTCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACGCTCCCGAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCCAAGTATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTCAACCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACCTGATGTTCTGAAAACATCATCACTGAG ATCATCCACTCCATCCTGCTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCCTTGGGATTTCGAGGCAGCTGAGTCCTCACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGTGGAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 108	464 aa	MW at 50403.4kD
NOV21l, 174308327 Protein Sequence	KLPTAVLILGPKVIKEKLTOELKDNATSILOQLPLLSAMREKPAGGIPVLGSLVNTV LKHIIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGFNTPLVKTIIVEFHMTTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKL SFLVNALAKQVMNLLVPSLPNLV KNQLCPVIEASFNGMYADPLQLVKVPI SLSIDRLEFDLLYP AIKGDITQLYLGAKLLD SQGKVTWKFNNASAASLTMTPLDNIPFSLIVSQDVVKA AVAVLSPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDAPEFFIDQGHAKVAQLIVLEVFPSSS ALRPLFTLGI EASSEAQFYTKGDQLILNLNNISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSVGVPSLVKALGFEESSSLTKDALVLT PASLWKPSPPVSQLE		
	SEQ ID NO: 109	1392 bp	
NOV21m, 174308337 DNA Sequence	AAGCTTCCCACTGCAGTTCTCATCCTCGGCCAAAAGTCATCAAAGAAAAGCTGACAC AGGAGCTGAAGGACCACAACGCCACCAGCATCCTGCAGCAGCTGCCGCTGCTCAGTGC CATGCGGGAAAAGCCAGCCGAGGCATCCCTGTGCTGGGCAGCCTGGTGAACACCGTC CTGAAGCACGTCACTGGCTGAAGGTCAACAGCTAACATCCTCCAGCTGCAGGTGA AGCCCTCGGCCAATGACCAGGAGCTGCTAGTCAAGATCCCCCTGGACATGGTGGCTGG ATTCAACACGCCCTCGGCCAAGACCATCGTGGAGTTCCACATGACGACTGAGGCCCAA		



	GCCACCATCCGCATGGACACCAGTGCAGTGGCCCCACCCGCCTGGTCTCAGTGACT GTGCCACCAGCCATGGGAGCCTGCGCATCCAAGTCTGCTGCATAAGCTCTCCTTCTGGT GAACGCCTTAGCTAAGCAGGTGATGAACCTCCTAGTGCCATCCCTGCCCAATCTAGTG AAAAACCAGCTGTGTCCCGTGATCGAGGCTTCTTCAATGGCATGTATGCAGACCTCC TGCAGCTGGTGAAGGTGCCATTTCCCTCAGCATTGACCGTCTGGAGTTTGACCTTCT GCATCCTGCCATCAAGGTGACACCATTGAGTCTACCTGGGGGCCAAGTTGTTGGAC TCACAGGGAAGGTGACCAAGTGGTTCAATAACTCTGCAGCTTCCCTGACAATGCCCA CCCTGGACAACATCCCGTTTCAGCCTCATCGTGAGTCAGGACGTGGTGAAAGCTGCAGT GGCTGCTGTGCTCTCTCCAGAAGAATTCATGGTCTCTGTTGGAAGTCTGTGCTTCCCTGAG AGTGCCCATCGGCTGAAGTCAAGCATCGGGCTGATCAATGAAAAGGCTGCAGATAAGC TGGGATCTACCCAGATCGTGAAGATCCTAACTCAGGACACTCCCAAGTTTTTTATAGA CCAAGGCCATGCCAAGGTGGCCCAACTGATCGTGCTGGAAGTGTTTCCCTCCAGTGAA GCCCTCCGCCCTTTGTTTACCCTGGGCATCGAAGCCAGCTCGGAAGCTCAGTTTTACA CCAAAGGTGACCAACTTATACTCAACTTGAATAACATCAGCTCTGATCGGATCCAGCT GATGAACTCTGGGATTGGCTGGTTCCAACTGATGTTCTGAAAAACATCATCACTGAG ATCATCCACTCCATCCTACTGCCGAACCAGAATGGCAAATTAAGATCTGGGGTCCCAG TGTCATTGGTGAAGGCTTGGGATTGAGGCAGCTGAGTCTCTACTGACCAAGGATGC CCTTGTGCTTACTCCAGCCTCCTTGGGGAAACCCAGCTCTCCTGTCTCCAGCTCGAG		
	ORF Start: AAG at 1	ORF Stop:	
	SEQ ID NO: 110	464 aa	MW at 50251.2kD
NOV21m, 174308337 Protein Sequence	KLPTAVLILGPKVKEKLTQELKDHNATSILOQLPLLSAMREKPAAGGIPVLGSLVNTV LKHVIWLKVITANILQLQVKPSANDQELLVKIPLDMVAGNTPLAKTIVEFHMTEAQ ATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKLFLVNALAKQVNNLLVPSLNLV KNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEFDLLHPAIGKDTIQLYLGAKLLD SQGKVTKWFNNSAASLTMPDLNIPFSLIVSQDVVKAAVAALVSPPEEFMVLLDSVLPE SAHRLKSSIGLINEKAADKLGSTQIVKILTQDTPKFFIDQGHAKVAQLIVLEVPFSS ALRPLFTLGIEASSEAFYTKGDLILNLNLISSDRIQLMNSGIGWFQPDVLKNIITE IIHSILLPNQNGKLRSQVPSLVKALGFEEAESSLTKDALVLTASLGLKPPSPVSQL		
	SEQ ID NO: 111	1023 bp	
NOV21n, CG59446-02 DNA Sequence	CCTCTGACACCTGGGAAGATGGCCGGCCCGTGGACCTTACCCTTCTCTGTGGTTGC TGGCAGCCACCTTGATCCAAGCCACCCTCAGTCCCACTGCAGTTCTCTCATCTCGGCC AAAAGTCATCAAAGAAAAGCTGACACAGGAGCTGAAGGACCACAACGCCACCAGCATC CTGCAGCAGCTGCCGCTGCTCAGTGCCATGCGGGAAGAACAGCCAGCCGAGGCATCCCTG TGCTGGGCAGCCTGGTGAACACCGTCTCTGAAGCAGCTCATCTGGCTGAAGGTGATCAC AGCTAACATCCTCCAGCTGCAGGTGAAGCCCTCGGCCAATGACCAGGAGCTGCTAGTC AAGATCCCCCTGGACATGGTGGCTGGATTCAACACGCCCTTGGTCAAGACCATCGTGG AGTTCCACATGACGACTGAGGCCCAAGCCACCATCCGATGGACACCAAGTGCAGTGG CCCCACCGCCTGGTCTCAGTGACTGTGCCACCAGCCATGGGAGCTGCGCATCCAA CTGCTGCATAAGCTCTCCTTCTGGTGAACGCCCTTAGCTAAGCAGGTGATGAACCTCC TAGTGCCATCCCTGCCCAATCTAGTGAAAAACAGCTGTGTCCCGTGATCGAGGCTTC CTTCAATGGCATGTATGCAGACCTCCTGCAGCTGGTGAAGGTGCCCATTTCCCTCAGC ATTGACCGTCTGGAGTTTGACCTTCTGTATCCTGCCATCAAGGTGACACCATTGAGC TCTACCTGGGGGCCAAGTTGTTGGACTCACAGGGAAGGTGACCAAGTGGTTCAATAA CTCTGCAGCTTCCCTGACAATGCCACCCTGGACAACATCCCGTTGAGCCTCATCGTG AGTCAGGACGTGGTGAAAGCTGCAGTGGCTGCTGTGCTCTCTCCAGAAGAATTCATGG TCCTGTTGGACTCTGTGGTAAACCTCAGCACAGGCAGAGAATAGGGCCGCCAGGCC ACATCATAGGAATTTCTGAACACAGGGTGGCCCTAA		
	ORF Start: ATG at 19	ORF Stop: TAA at 1021	
	SEQ ID NO: 112	334 aa	MW at 36309.5kD
NOV21n, CG59446-02 Protein Sequence	MAGPWTFTLLCGLLAATLIQATLSPTAVLILGPKVKEKLTQELKDHNATSILOQLPL LSAMREKPAAGGIPVLGSLVNTVLKHVIWLKVITANILQLQVKPSANDQELLVKIPLDM VAGFNTPLVKTIVEFHMTEAQATIRMDTSASGPTRLVLSDCATSHGSLRIQLLHKL FLVNALAKQVNNLLVPSLNLVKNQLCPVIEASFNGMYADLLQLVKVPISLSIDRLEF DLLYPKIGDTIQLYLGAKLLDSQGKVTKWFNNSAASLTMPDLNIPFSLIVSQDVVK AAVAALVSPPEEFMVLLDSVNNLSTRQIRIGPPRPHRNFLNTGCP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 21B.

<b>Table 21B. Comparison of NOV21a against NOV21b through NOV21n.</b>		
<b>Protein Sequence</b>	<b>NOV21a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV21b	25..439 3..462	388/472 (82%) 393/472 (83%)
NOV21c	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21d	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21e	25..439 3..462	404/468 (86%) 404/468 (86%)
NOV21f	25..439 3..462	405/468 (86%) 406/468 (86%)
NOV21g	25..439 3..462	405/468 (86%) 406/468 (86%)
NOV21h	25..439 3..462	404/468 (86%) 406/468 (86%)
NOV21i	25..439 3..462	403/468 (86%) 404/468 (86%)
NOV21j	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21k	25..439 3..462	406/468 (86%) 406/468 (86%)
NOV21l	25..439 3..462	405/468 (86%) 405/468 (86%)
NOV21m	25..439 3..462	402/468 (85%) 404/468 (85%)
NOV21n	1..318 1..310	308/318 (96%) 310/318 (96%)

Further analysis of the NOV21a protein yielded the following properties shown in Table 21C.

<b>Table 21C. Protein Sequence Properties NOV21a</b>	
<b>PSort analysis:</b>	0.6138 probability located in outside; 0.4772 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)

SignalP analysis:	Likely cleavage site between residues 25 and 26
-------------------	---

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 21D.

Table 21D. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY77126	Human neurotransmission-associated protein (NTAP) 2799056 - Homo sapiens, 484 aa. [WO200001821-A2, 13-JAN-2000]	1..439 1..484	431/492 (87%) 431/492 (87%)	0.0
AAG63976	Amino acid sequence of a human Lng103 polypeptide - Homo sapiens, 484 aa. [WO200161055-A2, 23-AUG-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAU29163	Human PRO polypeptide sequence #140 - Homo sapiens, 484 aa. [WO200168848-A2, 20-SEP-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAB87564	Human PRO1357 - Homo sapiens, 484 aa. [WO200116318-A2, 08-MAR-2001]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0
AAB66124	Protein of the invention #36 - Unidentified, 484 aa. [WO200078961-A1, 28-DEC-2000]	1..439 1..484	430/492 (87%) 431/492 (87%)	0.0

- In a BLAST search of public sequence databases, the NOV21a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 21E.

Table 21E. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96HK6	SIMILAR TO DNA SEGMENT, CHR 2, MASSACHUSETTS INSTITUTE OF TECHNOLOGY 19 - Homo sapiens (Human), 484 aa.	1..439 1..484	428/492 (86%) 431/492 (86%)	0.0
Q61114				e-127

	GLAND PROTEIN - Mus musculus (Mouse), 474 aa.	1..473	324/482 (66%)	
Q9BWZ6	DJ1187J4.1.1 (NOVEL PROTEIN SIMILAR TO MOUSE VON EBNER SALIVARY GLAND PROTEIN, ISOFORM 1.) - Homo sapiens (Human), 285 aa (fragment).	200..439 1..285	232/293 (79%) 232/293 (79%)	e-116
Q9BQP8	BA49G10.6 (SIMILAR TO MURINE VON EBNER MINOR SALIVARY GLAND PROTEIN, ISOFORM 1) - Homo sapiens (Human), 199 aa (fragment).	1..199 1..199	199/199 (100%) 199/199 (100%)	e-107
Q9H4V6	DJ1187J4.1.2 (NOVEL PROTEIN SIMILAR TO MOUSE VON EBNER SALIVARY GLAND PROTEIN, ISOFORM 2.) - Homo sapiens (Human), 213 aa.	272..439 1..213	160/221 (72%) 160/221 (72%)	1e-73

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21F.

Table 21F. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

#### EXAMPLE 22.

The NOV22 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis		
	SEQ ID NO: 113	2020 bp
NOV22a, CG59375-01 DNA Sequence	CATGAGTGAATGAAGGCACATGACAAACCTCCAGACCTGTGGAGACTGAAGGCTGAGA GCCTTTATAGATGCTGTGGGGCCGAGGAGTTTGCCAACTACAGCAGGTCATGCCAGC GCTGCAAGAGGCTACGTGCGGGTGGTCACCGAGAAGTCCCGACCGACTGGGCTCTC TTTACCTATGAAGGCAACAGCAATGACATCCGCGTGGCTGGCACAGGGGAGGGTGGCC TGGAGGAGATGGTGGAGGAGCTCAACAGCGGGAAGGTGATGTACGCCTTCTGCAGAGT GAAGGACCCCAACTCTGGACTGCCCAAATTTGTCTCATCACTGGACAGGCGAGGGC GTGAACGATGTGCGGAAGGGAGCCTGTGCCAGCCACGTCAGCACCATGGCCAGCTTCC TGAAGGGGGCCCATGTGACCATCAACGCACGGGCCGAGGAGGATGTGGAGCCTGAGTG CATCATGGAGAAGGTGGCCAAAGGCTTCAGGTGCCAACTACAGCTTTCACAAGGAGAGT GGCCGCTTCCAGGACGTGGGACCCAGGCCCCAGTGGTGAGTGGCTCTGTGTACCAGA AGACCAATGCCGTGTCTGAGATTAAAGGGTTGGTAAAGACAGCTTCTGGGCCAAAGC AGAGGACCTTGAGACCTTGTGAGAAAGAAATAAAGAGAAAGAGAGGAGGAGGCACAG CGGCAGCTGGAGCAGGAGCGCCGGGAGCGTGAGCTGCGTGAGGCTGCACGCCGAGAGC AGCGCTATCAGGAGCAGAGGTGGCGAGGCCAGAGCAGGACGTGGGAGCAGCAGCAAGA AGTGGTTTCAAGGAACCGAAATGAGCAGGGGTCAACATGTGCTTCCCTCCAGGAGTCT	

	GCCGTGCACCCGAGGGAGATTTTCAAGCAGAAGGAGAGGGCCATGTCTCACCACCTCCA TCTCCAGTCTCTCAGCCTGGCAAGCTGAGGAGCCCTTCTCTGCAGAAGCAGCTCACCCA ACCAGAGACCCACTTTGGCAGAGAGCCAGCTGCTGCCATCTCAAGGCCCAGGGCAGAT CTCCCTGCTGAGGAGCCGGCGCCAGCACTCCTCCATGTCTGGTGCAGGCAGAAGAGG AGGCTGTGTATGAGGAACCTCCAGAGCAGGAGACCTTCTACGAGCAGCCCCACTGGT GCAGCAGCAAGGTGCTGGCTCTGAGCACATTGACCACCACATTAGGGGCCAGGGGCTC AGTGGGCAAGGGCTCTGTGCCGTGCCCTGTACGACTACCAGGCAGCCGACGACACAG AGATCTCCTTTGACCCCGAGAACCTCATCACGGGCATCGAGGTGATCGACGAAGGCTG GTGGCGTGGCTATGGGCCGGATGGCCATTTTGCATGTTCCCTGCCAACTACGTGGAGC TCATTGAGTGAGGCTGAGGGGCACATCTTGCCCTTCCCCTCTCAGACATGGCTTCTCTTA TTGCTGGAAGAGGAGGCCTGGGAGTTGACATTGAGCACTCTTCCAGGAATAGGACCCC CAGTGAGGATGAGGCCTCAGGGCTCCCTCCGGCTTGGCAGACTCAGCCTGTCACCCCA AATGCAGCAATGGCCTGGTGATTCCACACATCCTTCTCTGCATCCCCCGACCTCCCA GACAGCTTGGCTCTTGCCCTGACAGGATACTGAGCCAAGCCCTGCCTGTGGCCAAGC CCTGAGTGGCCACTGCCAAGCTGCGGGGAAGGGTCTCTGAGCAGGGGCATCTGGGAGGC TCTGGCTGCCTTCTGCATTATTTGCCTTTTTTCTTTTTTCTTCTGCTTCTAAGGGGTG GTGGCCACCACTGTTTAGAATGACCCTTGGGAACAGTGAACGTAGAGAATNGTTTTTA GCAGAGTTGTGACCAAAGTCAGAGTGGATCATGGTGGTTTGGCAGCAGGGAATCTGTC TTGTTGGAGCCTGCTCTGTGCTCCCCACTCCATTCTCTGTCCCTCTGCCTGGGCTAT GGGAAGTGGGGATGCAGATGGCAAGCTCCCACCTGGGTATTCAAAA		
	ORF Start: ATG at 10	ORF Stop: TGA at 1585	
	SEQ ID NO: 114	525 aa	MW at 58507.2kD
NOV22a, CG59375-01 Protein Sequence	MKAHDKPPDLWRLKAESLYRCCGAEEFANYSRSCPALQEAAYVRVVTESPTDWALFTY EGNSNDIRVAGTGEGGLEEMVEELNSGKVMYAFRCRVKDPNSGLPKFVLINWTGEGVND VRKGACASHVSTMASFLKGAHVTINARAEDVEPECIMEKVAKASGANYSFHKEGRF QDVGPQAPVVSQSVYQKTNAVSEIKRVGKDSFWAKAEDPETLSERNKREREEEAQRQL EQERRERELREARREQRYQQRWRGQSRTWEQQQEVVSRNRNEQQGSTCASLQESAVH PREIFKQKERAMSTTSSISSPQPKLRSPFLQQLTQPEHFGREPAAAI SRPRADLPA EPPAPSTPPCLVQAEAEAVYEEPPEQETFYEQPPPLVQQQAGSEHIDHHIQGQGLSGC GLCARALYDYQAADDTEISFDPENLTGIEVIDEGWWRGYGPDGHFACSLPTTWSSLS EAEGTSCPSPLRHGFLIAGRGLGVDIQHSSRNRTPEDEASGLPPAWQTQPVTPNAA MAW		

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

- A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/	Identities/ Similarities for	Expect Value

		Residues	Region	
AAB93895	Human protein sequence SEQ ID NO:13840 - Homo sapiens, 439 aa. [EP1074617-A2, 07-FEB-2001]	28..465 3..439	407/440 (92%) 411/440 (92%)	0.0
AAY85662	Human tyrosine kinase substrate tks118/Dresh protein sequence - Homo sapiens, 431 aa. [WO200061750-A2, 19-OCT-2000]	28..465 3..431	399/440 (90%) 403/440 (90%)	0.0
AAB20896	Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1 - Homo sapiens, 431 aa. [JP2000197489-A, 18-JUL-2000]	28..465 3..431	398/440 (90%) 403/440 (91%)	0.0
AAM79569	Human protein SEQ ID NO 3215 - Homo sapiens, 458 aa. [WO200157190-A2, 09-AUG-2001]	28..465 31..458	397/439 (90%) 401/439 (90%)	0.0
AAM78585	Human protein SEQ ID NO 1247 - Homo sapiens, 430 aa. [WO200157190-A2, 09-AUG-2001]	28..465 3..430	397/439 (90%) 401/439 (90%)	0.0

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96K74	CDNA FLJ14461 FIS, CLONE MAMMA1000173, HIGHLY SIMILAR TO HOMO SAPIENS SRC HOMOLOG Y 3 DOMAIN-CONTAINING PROTEIN HIP-55 MRNA - Homo sapiens (Human), 439 aa.	28..465 3..439	407/440 (92%) 411/440 (92%)	0.0
Q96F30	SIMILAR TO SRC HOMOLOG Y 3 DOMAIN-CONTAINING PROTEIN HIP-55 - Homo sapiens (Human), 431 aa.	28..465 3..431	399/440 (90%) 403/440 (90%)	0.0
Q9UJU6	SRC HOMOLOG Y 3 DOMAIN-CONTAINING PROTEIN HIP-55 (DREBRIN F) - Homo sapiens (Human), 430 aa.	28..465 3..430	397/439 (90%) 401/439 (90%)	0.0
Q9NR72	CERVICAL SH3P7 (MUCIN-	28..465 3..430	395/439 (89%) 401/439 (90%)	0.0

	sapiens (Human), 430 aa.			
Q62418	DREBRIN-LIKE SH3 DOMAIN-CONTAINING PROTEIN SH3P7 - Mus musculus (Mouse), 433 aa.	29..465 4..433	345/439 (78%) 371/439 (83%)	0.0

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cofilin_ADF: domain 1 of 1	35..158	27/151 (18%) 101/151 (67%)	7.8e-21
SH3: domain 1 of 1	408..455	16/58 (28%) 31/58 (53%)	0.0038
Peptidase_M36: domain 1 of 1	486..509	11/24 (46%) 13/24 (54%)	2.9

#### EXAMPLE 23.

The NOV23 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis		
	SEQ ID NO: 115	2898 bp
NOV23a, CG59321-01 DNA Sequence	CTCTGGTCACAACCTGCATCAATGACATGCAGAAAGCAGGT CAGGT CACAAGATGCAGC CCTTTCCCGAGCTCTCTCGGAGAGTTGGATCTGATGGCAGGAACAGCAATGGCGAA GCCCTTCCCGAATCCATCCCATCAGCTCCTGGGACACTGCCTCATTTCATAGAGGAGC CAGATGATGCTTATATTATCAAGAGCAACCTATTGCACACTCAGGTGCAAAGCGAGGCC AGCCATGCAGATATTCTTCAAATGCAACGGCGAGTGGGTCCATCAGAACGAGCACGTC TCTGAAGAGACTCTGGACGAGAGCTCAGGTTTGAAGGTCCGCGAAGTGTTCATCAATG TTACTAGGCAACAGGTGGAGGACTTCCATGGGCCCCGAGGACTATTGGTGCCAGTGTGT GGCGTGGAGCCACCTGGGTACCTCCAAGAGCAGGAAGGCCTCTGTGCGCATAGCCTAT TTACGGAAAACTTTGAACAAGACCACAAAGGAAGGGAAGTTCCCATTTGAAGGCATGA TTGTACTGCACTGCCGCCACCAGAGGGAGTCCCTGCTGCCGAGGTGGAATGGCTGAA AAATGAAGAGCCATTGACTCTGAACAAGACGAGAATGACACCCAGGGCTGACCAT AACCTGATCATCAGGCAGGCACGGCTCTCGGACTCAGGAAATTACACCTGCATGGCAG CCAACATCGTGGCTAAGAGGAGAAGCCTGTCCGCCACTGTTGTGGTCTACGTGAATGG AGGCTGGTCTTCTGGACAGAGTGGTCAGCCTGCAATGTTGCTGTGGTAGAGGATGG CAGAAACGTTCCCGGACCTGCACCAACCCAGCTCCTCTCAATGGTGGGCGCTTTGTG AGGGAATGTGAGTGCAGAAAATAACCTGCACTTCTCTTTGTCTGTGGATGGGAGCTG GGAAGTGTGGAGCGAATGGTCCGTCTGCAGTCCAGAGTGTGAACATTGCGGATCCGG GAGTGCACAGCACACCCCCGAGAAATGGGGGCAAATTCTGTGAAGGTCTAAGCCAGG AATCTGAAAACATGCACAGATGGTCTTTGCATCCTAACTCCACCACCATGCAGGAACC CAAGGTGACTGCCCTTCAGACGCTATGCCAAATTGAGAATGCCAGCGACATTGCTTTG TACTCGGGCTTGGGTGCTGCCGTCTGGCCGTTGCAGTCTGGTCATTGGTGTCAACC TTTACAGACGGAGCCAGAGTACTATGGCGTGGACGTATTGACTCTTCTGCATTGAC AGGTGGCTTCCAGACCTTCAACTTCAAAACAGTCCGTCAAGGGAACCTCCCTGCTCCTG	

	AATTCTGCCATGCGACCCAGATCTGACAGTGAGCCGGACATACAGCGGACCCATCTGCTG TGCAGGACCCCTCTGGACAAGGAGCTCATGACAGAGTCCCTACTCTTTAACCCCTTTGTC GGACATCAAAGTGAAAGTCCAGAGCTCGTTCATGGTTTCCCTGGGAGTGTCTGAGAGA GCTGAGTACCACGGCAAGAATCATTCCAGGACTTTTCCCATGGAAACAACCACAGCT TTAGTACAATGCATCCCAGAAATAAAATGCCCTACATCCAAAATCTGTCACTACTCCC CACAAGGACAGAAGTGGAGACAAGTGGTGTCTTTGGCCATTTAGGGGGGCGCTTAGTA ATGCCAAATACAGGGGTGAGCTTACTCATACCACACGGTGCCATCCCAGAGGAAATT CTTGGGAGATTATATGTCCATCAACCAAGGTGAACCCAGGTGAGATGGCTCTGAGGT GCTCCTGAGTCTCTGAAGTCACTGTGGTCTCCAGACATGATCGTCACCACTCCCTTT GCATTGACCATCCCGCACTGTGCAGATGTGAGTTCTGAGCATTGGAATATCCATTTAA AGAAGAGGACACAGCAGGGCAAATGGGAGGAAGTGATGTGAGTGAAGATGAATCTA ATCCTGTTACTGCCTTTTGGACCCCTTTGCGTGTGATGTGCTCCTGGACAGCTTTGGG ACCTATGCGCTCACTGGAGAGCCAATCACAGACTGTGCCGTGAAGCAACTGAAGGTGG CGGTTTTTGGCTGCATGTCTGTAACTCCCTGGATTACAACTTGAGAGTTTACTGTGT GGACAATACCCCTTGTGCATTTCAAGGAAGTGGTTTCAGATGAAAGGCATCAAGGTGGA CAGCTCCTGGAAGAACCAAAATGCTGCATTTCAAAGGGAATACCTTTAGTCTTCAGA TTTCTGTCTTGATATTTCCCCATTCTCTGGAGAATTAACCACTTCACTGCCTGCCA GGAAGTCCCGTCTCCCGCGTGTGGTGCAGTAACCGGCAGCCCTGCCATGTGCCTTC TCCCTGGAGCGTTATACGCCCCTACCAACCCAGCTGTCTGCAAAATCTGCATTCCGGC AGCTCAAAGGCCATGAACAGATCCTCCAAGTGCAGACATCAATCCTAGAGAGTGAACG AGAAACCATCACTTTCTTCGCACAAGAGGACAGCACTTTCCCTGCACAGACTGGCCCC AAAGCCTTCAAATTTCCCTACTCCATCAGACAGCGGATTTGTGCTACATTTGATACCC CCAATGCCAAAGGCAAGGACTGGCAGATGTTAGCACAGAAACAGCATCAACAGGAG GAATTTATCTTATTTCGTACACAAAGTAGCCCATCTGTGTGATTTGAACCTGTGG GAAGCTCGTCATCAGCATGATGGTGTCTTGACTCCCTGGCCTGTGCCCTTGAAGAGA TTGGGAGGACACACAGAACTCTCAAACATTTCAGAATCCAGCTTGATGAAGCCGA CTTCAACTACAGCAGGCAAAATGGACTCTAGTCCACTTCTCCCATGAGACAGAGT		
	ORF Start: ATG at 21	ORF Stop: TAG at 2871	
	SEQ ID NO: 116	950 aa	MW at 105960.6kD
NOV23a, CG59321-01 Protein Sequence	MTCRKQVRSQDAALSQTLFGLDLDMAGTDNGEALPESIPSAPGTLPHFIEEPDDAYII KSNPIALRCKARPAMQIFPKNGEWHQNEHVSEETLDESSGLKVREVFINVTRQQVE DFHGPEDYWCQVAVSHLGTSSKSRKASVRIAYLRKNFEQDPQGREVP IEGMIVLHCRP PEGVPAAEVEWLKNEEPIDSEQDENIDTRADHNLII RQARLSDSGNYTCMAANIVAKR RSLSATVVVVYVNGGWSSWTEWSACNVRCGRGWQKRSRTCTNPAPLNGGAFCEGMSVQK ITCTSLCPVDGSWEVWSEWSVCSPECEHLRIRECTAPPPRNGGKFCEGLSQESENCTD GLCILNSTTMQEPKVTALQTLQIENASDIALYSGLGAAVVAVALVIGVTLYRRSQS DYGVDVIDSSALTGGFQTFNFKTVRQGNLLLLNSAMQPDLTVSRTYSGPICLQDPLDK ELMTESSLFNPLSDIKVKVQSSFMVSLGVSERAEYHGKNHSRTFPHGNHNSFSTMHPR NKMPYIQNLSSLPTRELRTTGVFGHLGGRLVMPNTGVSLLI PHGAI PEENSWEIYMS INQGEPRSDGSEVLLSPEVTCGPPDMIVTTPFALTI PHCADVSSEHWNHLLKRTQQG KWEVMSVEDESTCYCLLDPFACHVLLDSFGTYALTGEPITDCAVKQLKVAVFGCMS CNSLDYNLRVYCDNTPCAFQEVVSDERHQGGQLLEPKLLHFKGNTFSLQISVLDIP PFLWRIKPFTACQEVPPFSRVWCNSNRQPLHCAFSLERYTPTTQLSCKICIRQLKGHEQ ILQVQTSILESERETITFFAQEDSTFPAQTGPKAFKIPYSIRQRICATFDTPNAKGKD WQMLAQKNSINRRNLSYFATQSSPSAVILNLWEARHQHDGDLDSLACALEEIGRTHTK LSNISESQLDEADFNYSRQNGL		
	SEQ ID NO: 117	2181 bp	
NOV23b, CG59321-02 DNA Sequence	CGGCCAGTCAGAACAACTCCTCCTGTTTTTAATGAATTGGGTTTACCATTGACAAATGCT TCCTGATTTTCGTTGTTGACTTAAGCATGAATAGTAAGAGGCTCTGGTCACAAC TGCA TCAATGACATGCAGAAAGCAGGTGCGCGCGCTGGCTCCCGTGGCTGGGGCTGTGTTTC TGGGCGGGAGGGAACGGGGTGGCCCAAGGAAGTGAACATGGCGAAGCCCTTCCCGAA TCCATCCCATCAGCTCCTGGGACACTGCCTCATTTATAGAGGAGCCAGATGATGCTT ATATTATCAAGAGCAACCTATTGCACTCAGGTGCAAAGCGAGGCCAGCCATGCAGAT ATTCTTCAAATGCAACGGCGAGTGGGTCCATCAGAACGAGCACGTCTCTGAAGAGACT CTGGACGAGAGCTCAGGTTTGAAGTCCGCGAAGTGTTTCAATGTTACTAGGCAAC AGGTGGAGGACTTCCATGGGCCGAGGACTATTGGTGCCAGTGTGTGGCGTGGAGCCA CCTGGGTACCTCCAAGAGCAGGAAGGCTCTGTGCGCATAGCCTATTACGGA AAAAC TTTGAACAAGACCCACAAGGAAGGAAGTTCCTTGAAGGCATGATTGTACTGCACT		



	GCGCCCCACCAGAGGGAGTCCCTGCTGCCGAGGTGGAATGGCTGAAAAATGAAGAGCC CATTGACTCTGAACAAGACGAGAACATTGACACCAGGGCTGACCATAACCTGATCATC AGGCAGGCACGGCTCTCGGACTCAGGAAATTACACCTGCATGGCAGCCAACATCGTGG CTAAGAGGAGAAGCCTGTGCGCCACTGTTGTGGTCTACGTGAATGGAGGCTGGTCTTC CTGGACAGAGTGGTCAGCCTGCAATGTTGCTGTGGTAGAGGATGGCAGAAACGTTCC CGGACCTGCACCAACCCAGCTCCTCTCAATGGTGGGGCCTTTGTGAGGGAATGTCAG TGCAGAAAATAACCTGCACCTCTCTTTGTCTGTGGATGGGAGCTGGGAAGTGTGGAG CGAATGGTCCGTCTGCAGTCCAGAGTGTGAACATTTGCGGATCCGGGAGTGCACAGCA CCACCCCGAGAAATGGGGGCAAATTCTGTGAAGGTCTAAGCCAGGAATCTGAAAAC GCACAGATGGTCTTTGCATCTAGATAAAAAACCTCTTCATGAAATAAAACCCCAAAG CATTGAGAAATGCCAGCGACATTGCTTTGTACTCGGGCTTGGGTGCTGCCGTCTGTGGCC GTTGCACTCCTGGTCAATTGGTGTACCCCTTTACAGACGGAGCCAGAGTGACTATGGCG TGGACGTCATTGACTCTTCTGCATTGACAGGTAACCTCCCTGCTCCTGAATGCGAGCAC ACTCCAGCCTCTGGAGAGACGACAACGCGTGAAGCAACTGAAGGTGGCGGGTTTTGGC TGCATGTCTGTAACCTCCCTGGATTACAACCTGGAGAGTTTACTGTGTGGACAAAACCC CTTGGGCTTTTCAGGAAGTGGTTTCAGATGAAAGGCATCAAGGGGGACAGCTCCTGGA AGAACCAAATTTGCTGCATTTCAAAGGGAATACCTTTAGTCTTCAGATTCTGTCTCTT GATATTCCCCCATTCCTCTGGAGAATTAAACCATTCAGTGCCTGCCAGGAAGTCCCGG TCTCCCGCGTGTGGTGCAGTAACCGGCAGCCCTGCACTGTGCCTTCTCCCTGGAGCG TTATACCCCACTACCACCCAGCTGTCTTGCAAATCTGCATTCGGCAGCTCAAAGGC CATGAACAGATCCTCCAAGTGCAGACATCAATCCTAGAGACTGGCCCCAAAGCCTTCA AAATTCCTACTCCATCAGACAGCGGATTTGTGCTACATTTGATACCCCAATGCCAA AGGCAAGGACTGGCAGATGTAGCACAGAAAAACAGCATCAACAGGAATTTATCTTAT TTGCTACACAAAGTAGCCCATCTGCTGTCAATTTGAACCTGTGGGAAGCTCGTCATC AGCATGATGGTGATCTTGACTCCCTGGCCTGTGCCCTTGAAGAGATTGGGAGGACACA CACGAACTCTCAACATTCAGAATCCAGCTTGATGAAGCCGACTTCAACTACAGC AGGCAAAATGGACTCTAGTCCACTTCCTCCCATGA		
	ORF Start: ATG at 125	ORF Stop: TAG at 2162	
	SEQ ID NO: 118	679 aa	MW at 75724.8kD
NOV23b, CG59321-02 Protein Sequence	MQKAGRAAGSRGWGCVSGREGTGVAQGTDNGEALPESIPAPGTLPHFIEEPDDAYII KSNPIALRCKARPAMQIFFKNGEWHQNEHVSEETLDESSGLKVREVFINVTRQQVE DFHGPEDYWCQCVAWSHLGTSKSRKASVRIAYLRKNFEQDPQGREVPLEGMIVLHCRP PEGVPAAEVEWLKNEEPIDSEQDENIDTRADHNLII RQARLSDSGNYTCMAANIVAKR RSLSATVVVVYVNGGWSSWTEWSACNVRCGRGWQKRSRTCTNPAPLNGGAFCEGMSVQK ITCTSLCPVDGSWEVWSEWSVCSPECEHLRIRECTAPPRNGGKFCEGLSQESENCTD GLCILDKKPLHEIKPQSIENASDIALYSGLGAAVVAVAVLVIGVTLYRRSQSDYGV DV IDSSALTGNSLLNASTLQPLERRQRVKQLKVAGFGCMSCNSLDYNWRVYCVDKTPWA FQEVVSDERHQGGQLLEPKLLHFKGNTFSLQISVLDIPPFLWRIKPFTACQEVVPSR VWCSNRQPLHCAFSLERYPTTTQLSCKICIRQLKGHEQILQVQTSILETGPKAFKIP YSIRQRICATFDTFNAKGKDWMQAQKNSINRNLSYFATQSSPSAVILNLWEARHQHD GDLDSLACALEEIGRTHTKLSNISESQLDEADFNYSRQNGL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 23B.

Table 23B. Comparison of NOV23a against NOV23b and NOV23c.		
Protein Sequence	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV23b	27..444	357/419 (85%)
	27..426	361/419 (85%)

Further analysis of the NOV23a protein yielded the following properties shown in Table 23C.

<b>Table 23C. Protein Sequence Properties NOV23a</b>	
<b>PSort analysis:</b>	0.8411 probability located in mitochondrial inner membrane; 0.7000 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2057 probability located in mitochondrial matrix space
<b>SignalP analysis:</b>	No Known Signal Sequence Predicted

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 23D.

<b>Table 23D. Geneseq Results for NOV23a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV23a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAU12244	Human PRO4326 polypeptide sequence - Homo sapiens, 945 aa. [WO200140466-A2, 07-JUN-2001]	26..925 26..933	499/915 (54%) 651/915 (70%)	0.0
AAW78900	Rat UNC-5 homologue UNC5H-2 - Rattus sp, 943 aa. [WO9837085-A1, 27-AUG-1998]	25..925 23..931	487/916 (53%) 648/916 (70%)	0.0
AAB50691	Human UNC5C protein SEQ ID NO:90 - Homo sapiens, 931 aa. [WO200073328-A2, 07-DEC-2000]	34..936 49..930	465/921 (50%) 621/921 (66%)	0.0
AAW78898	Rat UNC-5 homologue UNC5H-1 - Rattus sp, 898 aa. [WO9837085-A1, 27-AUG-1998]	26..936 23..897	417/921 (45%) 582/921 (62%)	0.0
AAM79128	Human protein SEQ ID NO 1790 - Homo sapiens, 943 aa. [WO200157190-A2, 09-AUG-2001]	24..936 31..942	422/955 (44%) 588/955 (61%)	0.0

- In a BLAST search of public sequence databases, the NOV23a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 23E.

<b>Table 23E. Public BLASTP Results for NOV23a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV23a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
BAB83663	KIAA1777 PROTEIN (UNC5H4) - Homo sapiens (Human), 948 aa.	27..950 30..948	906/926 (97%) 910/926 (97%)	0.0

O08722	TRANSMEMBRANE RECEPTOR UNC5H2 - Rattus norvegicus (Rat), 945 aa.	25..925 25..933	488/916 (53%) 649/916 (70%)	0.0
Q9D398	6330415E02RIK PROTEIN - Mus musculus (Mouse), 945 aa.	1..925 1..933	491/940 (52%) 656/940 (69%)	0.0
O08747	ROSTRAL CEREBELLAR MALFORMATION PROTEIN - Mus musculus (Mouse), 931 aa.	34..936 49..930	468/921 (50%) 622/921 (66%)	0.0
O95185	TRANSMEMBRANE RECEPTOR UNC5C - Homo sapiens (Human), 931 aa.	34..936 49..930	465/921 (50%) 621/921 (66%)	0.0

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23F.

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig: domain 1 of 1	165..225	16/63 (25%) 43/63 (68%)	5.2e-07
tsp_1: domain 1 of 2	248..297	23/54 (43%) 33/54 (61%)	1.6e-07
tsp_1: domain 2 of 2	304..351	23/54 (43%) 32/54 (59%)	0.0014
ZU5: domain 1 of 1	538..638	33/115 (29%) 68/115 (59%)	7.8e-21
death: domain 1 of 1	855..933	21/87 (24%) 61/87 (70%)	4.4e-13

#### EXAMPLE 24.

The NOV24 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis		
	SEQ ID NO: 119	898 bp
NOV24a, CG59591-01 DNA Sequence	CCTGTGACTCCTCCATCCAGCTATGCCCTGCTGCCAGCACCGTGGGCTGGCAGGC CTGCTCTTCTGGGCTGGCCAGGCAGTGAACGCCTTGATAATGCCTAATGCTACCCAG CCCCGGCCAGCCGAGAGCACGGCTATGCGGCTCCTGAGTGGCCTGGAGGTGCCAG GTACCGCCGGAAGCGCCACATCTCTGTGAGAGACATGAATGCCTTACTGGATTATCAC AACCACATCCGGCCAGTGTGTACCCACCTGCCGCCAACATGGAATACATGGTGTGGG ACAAGCGGCTGGCCAGGGCTGCCAAGCCTGGGCCACCCAGTGCATCTGGGCACATGG GCCTTCACAGCTGATGAGATACGTGGGCCAGAACCTCTCCATCCATTCTGGCCAGTAC CGGTCCGTAGTGGATCTCATGAAGTCTGGTCTGAGGAGAAGTGGCATTACTTGTTTC CGGCCCAAGGACTGTAACCCACACTGCCCTGGCGCTGCGATGGCCCCACCTGCTC	

	CCATTATACCCAGATGGTGTGGGCATCCTCCAATCGGCTGGGCTGTGCCATCCACACC TGTAGTAGCATCAGTGTCTGGGGCAACACCTGGCATCGGGCGGCATACCTGGTCTGCA ACTATGCCATTAAGGGCAACTGGATTGGCGAGTCCCCGTACAAGATGGGAAAGCCGTG CTCCTCTGTCCCCCAGTTATCAAGGCAGCTGCAATAGCAACATGTGCTTCAAGGGG CTGAAATCCAACAAGTTCACGTGGTTCTGAATTTCTCTGGGCTTTGGTGCGCCTCCA GCTGGGCCTGACCCTCCATGTCCTGCCCTCAAAAACTGGGTGGAGAAATAATTGTTT CTTTAAAGGATATGAGTTAGAATCACCC		
	ORF Start: ATG at 23	ORF Stop: TGA at 782	
	SEQ ID NO: 120	253 aa	MW at 28604.6kD
NOV24a, CG59591-01 Protein Sequence	MPLLPTVGLAGLLFWAGQAVNALIMP NATPAPAQPESTAMRLLSGLEVPYRRKRHI SVRDMNALLDYHNHIRASVYP PAA NMEYMWWDKRLARAAEAWATQCIWAHGPSQLMRY VGQNL SIHSGQYRSVVDLMKSWSEEKWHYLFPA PRDCNPHCPWRCDGPTCSHYTQMVW ASSNRLGCAIHTCSSISVWGN TWHRAYLV CNYAIKGNWIGESPYKMGKPCSSCPPSY QGSCNSNMCFKGLKSNKFTWF		

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

Table 24B. Protein Sequence Properties NOV24a	
PSort analysis:	0.4400 probability located in lysosome (lumen); 0.3798 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

- A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 24C.

Table 24C. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAR79914	Trypsin inhibitory protein, isolated from human T98G cells - Homo sapiens, 198 aa. [JP07242700-A, 19-SEP-1995]	57..253 1..197	130/197 (65%) 159/197 (79%)	3e-86
AAR79915	Human trypsin inhibitory protein, residues 11-198 - Homo sapiens, 188 aa. [JP07242700-A, 19-SEP-1995]	67..253 1..187	125/187 (66%) 152/187 (80%)	1e-83
AAU29058	Human PRO polypeptide sequence #35 - Homo sapiens, 500 aa. [WO200168848-A2, 20-SEP-2001]	19..243 10..242	112/235 (47%) 155/235 (65%)	3e-70
AAM41693	Human polypeptide SEQ ID NO 6624	19..243 93..325	112/235 (47%) 155/235 (65%)	3e-70

	[WO200153312-A1, 26-JUL-2001]			
AAM39907	Human polypeptide SEQ ID NO 3052 - Homo sapiens, 300 aa. [WO200153312-A1, 26-JUL-2001]	19..243 10..242	112/235 (47%) 155/235 (65%)	3e-70

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H3Y0	DJ881L22.3 (NOVEL PROTEIN SIMILAR TO A TRYPSIN INHIBITOR) - Homo sapiens (Human), 253 aa.	1..253 1..253	253/253 (100%) 253/253 (100%)	e-159
O43692	25 KDA TRYPSIN INHIBITOR - Homo sapiens (Human), 258 aa.	37..253 41..257	137/217 (63%) 170/217 (78%)	6e-90
Q98ST6	SUGARCRISP - Gallus gallus (Chicken), 258 aa.	22..253 20..257	140/238 (58%) 179/238 (74%)	8e-90
Q99MM7	SUGARCRISP - Mus musculus (Mouse), 258 aa.	3..253 2..257	144/256 (56%) 186/256 (72%)	1e-89
Q98ST5	COCOACRISP - Gallus gallus (Chicken), 523 aa.	14..243 14..242	111/230 (48%) 158/230 (68%)	6e-71

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SCP: domain 1 of 1	67..215	54/173 (31%) 96/173 (55%)	2.9e-21

5

#### EXAMPLE 25.

The NOV25 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis		
	SEQ ID NO: 121	2345 bp
NOV25a,	AGGAGCCGCGATGTTCCCCCTTCGGGCCCTGTGGTTGGTCTGGGCGCTTCTAGGAGTG	

CG59588-01 DNA Sequence	GCCGGATCATGCCCGGAGCCGTGCGCCTGCGTGGACAAGTACGCTCACCAGTTCGCGG ACTGCGCTTACAAAGAGTTGCGTGAGGTGCCGAAGGACTGCCCTGCCAAGTACGAC GCTTAGTCTGTCCGCGAACAAGATCACTGTGCTGCGGCGCGGGGCTTCGCGGACGTC ACACAGGTCACGTGCTGTGGCTGGCGCACAATGAGGTGCGCACCCTGGAGCCAGGCG CACTGGCCGTGTGAGTCAAGTCAAGAACCTCGATCTGAGCCACAACCTTCATATCCAG CTTTCGCTGGAGCGACCTGCGCAACCTGAGCGCGCTGCAGCTGCTCAAAATGAACCAC AACCGCCTGGGCTCTCTGCCCCGGGACGCACTCGGTGCGCTACCCGACCTGCGTTCCC TGCGCATCAACAACAACCGGCTGCGTACGCTGGCGCCTGGCACCTTCGACGCGCTTAG CGCGCTGTCACACTTGCAACTCTATCACAATCCCTTCCACTGCGGCTGCGGCTTGTG TGGCTGCAAGGCTGGGCGCGAGCACCCTGGGTGTCTTACCCGAGCCGACTCCATTG CTTGTGCTCGCTCCCGCGCTGCAAGGGGTGCCGGTGTACCGCTGCCCGCCCTGCC CTGTGCAACCGCCAGCGTGCATCTGAGTGCCGAGCCACCCTTGAAGCACCAGCCACC CCACTGCGCGCAGGACTGCGTTTCTGTTTACACTGCATCGCCGACGCCACCTTACGC CTCGCTGCAATGGCAACTTCAGATCCCCGGTGGCACCGTAGTCTTAGAGCCACCAGGT TCTGAGCGGGGAGGACGAGGGGTGGGGCGGAGGAAGGAGAGGAGAGAGAGTGGG GATTTGCTGACGCGAGACCAAGCCCAACGCGGACTCCAGCACCCTTGGCCGCGC CCCCAGCCACACCGGCTTCTGGCCCTCGCAAATGGCTCCCTGTTGGTGCCCTCCT GAGTGCCAAGGAGGCGGGCTCTACACTTGCCGTGCACACAATGAGTGGGCGGCCAAC TCTACGTCATACGCGTGGCGGTGGCAGCAACCGGGCCCCCAAAACAGCGCCTGGCG CCGGGGGAGAACCCGACGAGCAGGCCCGGACCTCTGAGCGCAAGTCCACAGCCAAGGG CCGGGGCAACAGCGTCTGCTTCCAAACCCGAGGGCAAAATCAAAGGCCAAGGCTG GCCAAGGTCAGCATTCTCGGGGAGACCGAGACGAGCCGAGGAGGAGACAAAGTGAAG GAGAGGAGGCCGAAGACAGATCCTCGCGGACCCGGCGGAGGAGCAGCGCTGTGGCAA CGGGGACCCCTCTCGGTACGTTTCTAACCACGCGTTCAACCAGAGCGCAGAGTCAAG CCGCACGTCTTCGAGCTGGGCGTCATCGCGCTGGATGTGGCGGAGCGGAGGCGCGGG TGCAGCTGACTCGCTGGCTGCGCGCTGGGGCCCTGGGCCCGCGGGGCTGGCGGAGC CCCGCGACCCGGGCGGCGACCCCTGCGCCTACTCTATCTGTGTCCAGCGGGGGCGGC GCGGCAGTGCAGTGGTCCCGCTAGAGGAAGGCGTCAACGCCTACTGGTTCGCGGCC TGCGGCCGGGTACCAACTACTCCGTGTGCTTGGCGCTGGCGGCGAGCGCTGCCAGT GCAAGTGGTGTTCACCAAGAAGGAGCTCCCATCGCTGCTGGTATAGTGGCAGTG AGCGTATTCCTCCTGGTGTGGCCACAGTGGCCCTTCTGGGCGCCGCTGCTGCCATC TGCTGGCTAAACACCCGGGCAAGCCCTACCGTCTGATCCTCGGCCTCAGGCCCCCTGA CCCTATGGAGAAGCGCATCGCCGCACTTCGACCCGCGTCTCGTACCTCGAGTCC GAGAAAAGTACCCGGCAGGCGGCGAGGCGGCGGCGAGGAGGAGGAGCTGCGAGG GGGAGGGCCTTGATGAAGACGCGGAGCAGGGAGACCAAGTGGGGACCTGCAGAGAGA GGAGAGCCTGGCGGCTGCTCACTGGTGGAGTCCAGTCCAAGGCCAACCAAGAGGAG TTCGAGCGGGCTCTGAGTACAGCGATCGGCTGCCCCTGGCGCCGAGGCGGTCAACA TCGCCAGGAGATTAATGGCAACTACAGGACAGCGGAGGCTGAACCTCCGCCCGTCC GGCCCGCCATCCCGACCTCCACCTAGGCTGCTGGGAGCAGCAGTCTAGGCTGGC AGGACTTATGTCCCCCGTCCCCAAC
	ORF Start: ATG at 11 ORF Stop: TGA at 2246
	SEQ ID NO: 122 2345 aa MW at 78989.2 kD
NOV25a, CG59588-01 Protein Sequence	MFPLRALWLVLWALLGVAGSCPEPCACVDKYAHQFADCAKELREVPEGLPANVTLSL SANKITVLRGAFADVTQVTSWLAHNEVRTVEPGALAVLSQLKNLDLSHNFISFPW SDLRNLALQLLKMNNRLGSLPRDALGALPDLRLRINNRLRLTLPAGTFDALSALS HLQLYHNPFFHCGGLVWLQAWAASRVSLPEPDSIACASPPALQGVVYRLPALPCAP PSVHLSAEPPLPAGTPLRAGLAFVLHCIADGHPTPRLQWLQIPGGTVVLEPPVLSG EDDGVAEEGEGEGDGLLTQTQAQTPTPAPAWPAPPATPRFLALANGSLLVPLLSAK EAGVYTCRAHNELGANSTSIRVAVAAATGPPKHAPGAGGEPDQAPTSEKSTAKGRGN SVLPSPKPEGKIKQGLAKVSI LGTETETPEEDTSEGEAEQILADPAEEQRCNGNDP SRYVSNHAFNQSAELKPHVFELGVIALDVAEREARVQLTPLAARWGPFGGAGGAPRP GRRPLRLLYLCPAGGGAQVQWSRVEGVNAYWFRGLRPGTNYSVCLALAGEACHVQVV FSTKKELPSLLVIVAVSVFLLVLATVPLLGAAACCHLLAKHPGKPYRLILRPQAPDPME KRIAADFDPRASYLESEKSPAGGEAGGEEPEDVQGEGLDEDAEQGDPDGLQREESL AACSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEIINGNYRQTAG

Further analysis of the NOV25a protein yielded the following properties shown in Table 25B.

<b>Table 25B. Protein Sequence Properties NOV25a</b>	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 25C.

<b>Table 25C. Geneseq Results for NOV25a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAE09450	Human sbg34976IGBa protein #1 - Homo sapiens, 745 aa. [WO200160850-A1, 23-AUG-2001]	1..745 1..745	745/745 (100%) 745/745 (100%)	0.0
AAU12205	Human PRO4329 polypeptide sequence - Homo sapiens, 745 aa. [WO200140466-A2, 07-JUN-2001]	1..745 1..745	745/745 (100%) 745/745 (100%)	0.0
AAB40448	Human ORFX ORF212 polypeptide sequence SEQ ID NO:424 - Homo sapiens, 209 aa. [WO200058473-A2, 05-OCT-2000]	265..472 2..209	208/208 (100%) 208/208 (100%)	e-120
AAM93734	Human polypeptide, SEQ ID NO: 3699 - Homo sapiens, 428 aa. [EP1130094-A2, 05-SEP-2001]	1..417 1..385	217/426 (50%) 260/426 (60%)	e-107
AAU12317	Human PRO215 polypeptide sequence - Homo sapiens, 428 aa. [WO200140466-A2, 07-JUN-2001]	1..417 1..385	217/426 (50%) 260/426 (60%)	e-107

- In a BLAST search of public sequence databases, the NOV25a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 25D.

<b>Table 25D. Public BLASTP Results for NOV25a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV25a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9P263	KIAA1465 PROTEIN - Homo	104..745 1..642	642/642 (100%) 642/642 (100%)	0.0

	(fragment).			
O14498	ISLR PRECURSOR - Homo sapiens (Human), 428 aa.	1..417 1..385	217/426 (50%) 260/426 (60%)	e-106
BAA85972	ISLR PRECURSOR - Mus musculus (Mouse), 428 aa.	4..417 1..385	209/421 (49%) 258/421 (60%)	e-102
O88279	MEGF4 - Rattus norvegicus (Rat), 1531 aa.	20..246 734..933	77/232 (33%) 113/232 (48%)	1e-25
Q9WVB5	SLIT1 - Mus musculus (Mouse), 1531 aa.	20..246 734..933	77/232 (33%) 113/232 (48%)	1e-25

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25E.

Table 25E. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRRNT: domain 1 of 1	19..50	12/33 (36%) 20/33 (61%)	0.27
LRR: domain 1 of 5	52..75	7/25 (28%) 18/25 (72%)	1.4
LRR: domain 2 of 5	76..99	5/25 (20%) 18/25 (72%)	31
LRR: domain 3 of 5	100..123	11/25 (44%) 20/25 (80%)	0.0026
LRR: domain 4 of 5	124..147	10/25 (40%) 18/25 (72%)	0.099
LRR: domain 5 of 5	148..171	9/25 (36%) 19/25 (76%)	0.14
LRRCT: domain 1 of 1	181..231	19/54 (35%) 41/54 (76%)	1.5e-14
ig: domain 1 of 1	253..357	13/108 (12%) 70/108 (65%)	0.01

#### EXAMPLE 26.

The NOV26 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis		
	SEQ ID NO: 123	4426 bp
NOV26a,	ATGTACTGTTTTTGCGATCTGCCGTTTCGTTCTTCTGCCTCAGCCTCCCCAGGTGCT	



CG59584-01 DNA Sequence	GGGGTTATAGGTGTGAGCCACTGTGCCTGGCTATTCTTTTATTACAGTATGTTCTGCT GATTCCCTCTGTTCTACAAGAAGGCTCTTTGGATAAAGCTTGTGCCAGCTTTTTAAT CTCACTGAATCTGTTGTTTTGACGGTCTCCCTCAACTATGGTGAGGTCCAGACCAAAA TATTTGAAGAAAATGTTACTGGAGAAAATTTCTCAAATGCATCAGCTTTGAGGTTCC TCAGGCCAGATCTGACCCACTGGCATTATTACATTTTCTGCTAAAGGAGCCACTCTC AACCTGGAAGAGAGGAGATCTGTGGCAATCAGATCCAGAGAGAATGTGGTCTTTGTAC AGACTGATAAACCCACCTACAAGCCTGGACAGTATAATAAAAGCCGATCAGTCACAT AATGCCAGTGATAGCAGTCACTGAACAGGATCCAGAAGGCAATCGAATACAACAGTGG GTGAATGAGGAGTCTGTGGGAGGGATTCTACAACCTCTCTTCCAGTTAATCTCAGAGC CCATCCTCGGATGGTATGAAATCACCGTGGAGATGCTCAATGAGAAGAAAACATATCA CTCCTTCTCTGTGGAAGAATATGTGTTACCCAAATTTCAAATGACTGTGGATGCACCA GAAAATATCTTAGTTGTGGACTCTGAATTCAAAGTGAATGTCTGTGCCTGTATACCT ATGGTGAACCTGTGGACGGGAAGGTCCAACCTTAGTGTGTGCAGAGAATCTACGGCTTA TCATTCACTGTGCTCATCTTATCAGTTCACTCTGTAAAAATTTTACCTTGGGGAAAGAT GGCTGTGTCTCCAAGTTTATTAACACAGATGCTTTTGAGTTAAATCGGGAAGGATACT GGAGTTTCTCAAGTGCATGCTCTTGTACAGAGCTTACAGGCTCCAAGTACGTATA CATAGACTCATCAGTGGTGAAGATTAGTTTGAAGAATATGGATATGCTCAACAACAG GGACTCCCTTATTTGGCCAGATTAAATGCTTAATCCAGACAACTCTCCAATCCCAA ATGAAGTTGTTCAAGTGCATCTGAAGGACAAAATCGTGGGAACTACACCACAGATGT AAATGGCATCGCTCAGTTTTTCTTGGACACATATACGTTTACATACCCAAATATCACT TTGAAAGCAGCCTACAAGGCAATGAAAATTTGCCAGGCTCATGGCTGGGTGTTGCCTC AATACCTCAGCCCGAGTACTTTGCATATCGATTTTACTCCAAGATGAATAGCTTCCT AAAGATTGTCCAAGAGATGGAAGAACTGAGATGCAACCAGCAGAAGAGGGTCTAGTG CACTGCATTCTCAATATGGAAGACTTTGAAGACAAAACCTACACAGCAGACTTCAATT ATTTGGTGATTTCAAAGGTGAATCATCTTCTATGGGCAACAGAAAATTTGAGATCAA CGAAAATGGGAGGAAGGGCATATTTTCCATTTCTATAGACATTAACCTGAATTAGCG CCCTCAGTACATATGCTTGTCTATAGCTTGCATCCTGGAGGAGAAAATGGTCACTGATA GCACCCAATTCCAATTGAGAAATGTTAACATAAAGTTCTCTAACGAGCAGGGCTTACC TGGTTCCAATGCTAGTCTCTGTCTTCAAGCGCGCTGTCTTATTCTGTGCCCTCAGG GCTGTGGATAGGAATGTCTTCTACTGAAATCTGAACAACAGCTGTGAGCTGAAAGTG TGTATAACATGGTTCCAAGTATAGAGCCGTATGGTTATTTCTACCATGGCTCAATCT TGATGATGGCAAGGAAGACCTTGCATTCTCAGAGGGATATGTTCTACAATGGTTTA TATTACACACCTGTAAGCAACTATGGGGATGGAGATATCTATAATTTGTCAGGAACA TGGGTCTAAAAGTCTTTACCAATCTCCATTACCGAAAACAGAAATATGTGTGATGGA GAGAAGGCTGCCACTCCCTAAGCCGCTTTATCTGGAAACAGAAAATTATGGTCCAATG CGTAGTGTCCGTCTAGAATTGCATCTAGTGGAATCAGAGGGGAGAATGCTGACTATG TAGAACAGGCTATAATTCAAACAGTAAGAACAACCTTCCAGAGACATGGATGTGGGA CCTCGTCAGTGTGATTCTCAGGCTCTGCCAATCTTTGTTCTCTCATTCTCTGATACG ATAACCCAATGGGAGGCAAGTGGCTTTTGTGTGAATGGTGACGTTGGATTGGCATT CCTCTACAACCACTCTAGAAGTCTCCCAACCTTTCTTTATTGAGATGCTCACCCTT TTGCGTTGTTCAAATGAACAATTTGATTTGATTGTCAATGTCTTCAGCTACCGGAAT ACATGTGTAGAGATTTCTGTTCAAGTGGAGGAGTCTCAGAATTATGAAGCAAATATTC ATACCTTGAATAATCAATGGCAGTGAGGTTATTCAAGCTGGAGGGAGGAAAACAAACGT CTGGACTATTATACCTAAGAAATTTGGGCAAAGTGAATATCACTGTAGTTGCTGAGTCC AAACAAAGCAGTGTCTGCCCAAATGAAGGAATGGAGCAGCAAAAGCTAACTGGAAG ACACTGTGGTCCAAAGCTTCTTAGTAGAGCCTGAAGGTATTGAAAAGGAAAGGACCCA GAGTTTCCTTATCTGTACAGAAGGTGCCAAAGCCTCCAAGCAGGGAGTTTGGACTTG CCAAACGATGTAGTAGAAGGGTCAGCCAGAGGCTTTTCACTGTTGTGGGGATATTC TAGGACTTGCCTTGCAATCTGGTTGTTCTCCAAATGCCCTATGGAAGTGGAGAGCA GAATGCTGCCCTACTAGCATCTGATACTTATGTTCTGGACTATCTGAAATCTACTGAG CAACTGACAGAGGAAGTTCAATCTAAGGCTTTCTTTCTTCTAATGTTATCAAAA GGCAATTATCTTCAAACCTCTGATGGTTCTATAGTGTGTTTTGGCAGCAGAGTCA GAAAGGAAGCATATGGCTCAGTGCTCTTACTTTTAAGACATTGGAGAGAATGAAAAAA TATGTATTCATTGATGAAAATGTTCAAACACAGACCTTAATCTGGCTTTCAAGCCAAC AGAAAACAAGCGGCTGCTTTAAGAAATGATGGCCAGCTTTTCAACCAGCCTGGCAGGG TGGAGATGAAGAGGACATTTCACTCACTGCGTATGTTGTTGGGATGTTCTTTGAAGCT GGGGCGGCATTGGACAGTGGTGTCACTAATGGCTATAATCATGCAATTTAGCTTATG CTTTTGCTTAGCTGGAAGAGAGAAGCAAGTGAATCTTTACTCCAACCTGGATCA ATCTGCCCAAACTAAATATGTCTACTGGGAAAGAGAAGGAAACCAAGACA GAAGAATTTCCATCCTTTATTCCTGGGCACCTTCTGCTCAGACTGAGAAGAGTTGCT ACGTGCTGTTGGCTGTCAATTCCCGGAAAATTCCTGACCTCAGCTATGCTAGTAAGAT
----------------------------	---

	TGTGCAGTGGCTTGCCCAACGGATGAATTCCCATGGAGGCTTTTCTTCCAACAGGAT CAAAACACTGTACCTTTAGCAGTGAAGGATCCAGTGAGATTTTCCAGGTTAACGGTC ATAACCGCTACTGGTCCAACGTTTCAGAAGTAACACAGGCACCTGGAGAATACACAGT AGATGTGGAAGGACACGGTTGTACATTTATCCAGGCCACCCTTAAGTACAATGTTCTC CTACCTAAGAAGGCATCTGGATTTTCTCTTCTTGGAAATAGTAAAGAACTACTCTT CGACTGCTTTTGACCTCACAGTGACCCTCAAATACACTGGAATTCGCAATAAATCCAG TATGGTGGTTATAGATGTAAAAATGCTATCAGGATTTACTCCAACCATGTATCCATT GAAGAGCTTGAAAAACAAGGCCAAGTGATGAAGACTGAAGTCAAGAATGACCATGTTT TTTTCTACTTGGAAAATGTAGGTTTTGGTCGAGCAGACAGTTTCCCTTTTTCTGTTGA GCAGAGCAACCTTGTGTTCAACATTCAGCCAGCCCCAGCCATGGTCTACGATTATTAT GAAAAAGAAGAATATGCCCTAGCTTTTACAACATCGACAGTAGTTTCAAGTTTCCGAGT GAGACAAAGCAATTACTAGAAGAGTTGGAGAAGCATTCTTGTAAACAACTGATTCTT CTGTATCAAACCTGGAAAAAATCATGAACCATCTGACATCGTGAACAGTCTGCAGTG GGCTATGGTTTTCTGTCAAGTCTTATTTCTTATCATCCATTAAATGTTGTCATTTT GCAAAAAAAAAAAAAA		
	ORF Start: ATG at 1	ORF Stop: TGA at 4234	
	SEQ ID NO: 124	1411 aa	MW at 158867.0kD
NOV26a, CG59584-01 Protein Sequence	MYCFLRSVAVFFCLSLPRCWGYRCEPLCLAILLQYVLLIPSVLQEGSLDKACAQLFN LTESVVLTVSLNYGEVQTKIFEENVGTGENFFKCSFEVPQARSDFLAFITFSAGKATL NLEERSVAIRSRENVVFQTDKPTYKPGQYNKKPISHIMPVIAVTEQDPEGNRQQW VNEESVGGILQLSFQLISEPILGWYEITVEMLNEKKTYHSFSVEEYVLPKFQMTVDAP ENILVVDSEFKVNVCALYTYGEPVDGKVQLSVCRESTAYHSCAHLISSLCKNFTLGKD GCVSKFINTDAFELNREGYWSFLKVHALVTELTGSKYVYIDSSVVKISFENMDMSYKQ GLPYFGQIKLLNPDNSPIPEVVQLHLKDKIVGNYTTDVNGIAQFFLDYTFYTPNIT LKAAYKANENCQAHGWLPQYPQPEYFAYRFYSKMNSFLKIVQEMEELRCNQKRVLV HCILNMFEDFKTYTADFNVLVISKGVILHGGQKIEINENGRKGIFISIDINPELA PSVHMLVYSLHPGGEMVTDSTQFQLRNVNIKFSNEQGLPGSNASLCLQAPVLFALR AVDRNVLLKSEQLSAESVYNMVPISIEPYGYFYHGLNLDDGKEDPCIPQRDMFYNGL YYTPVSNYGDGDIYNIVRNMGKLVFTNLHYRKPEVCVMERRLPLPKPLYLETENYGP RSVPSRIASSGIRGENADYVEQAIQIVRTNFPETWMDLVSDSSGSANLSFLIPDT ITQWEASGFCVNGDVGFGISSTTTLEVSQPFIEIASPFSVVQNEQFDLIVNVFSYRN TCVEISVQVEESQNYEANIHTLKINGSEVIQAGGRKTNVWTIIPKKLGKVNITVVAES KQSSACPNEGMEQKLNWKDVTVVQSFLVEPEGIEKERTQSFLICTEGAKASKQGVLDL PNDVVEGSARGFFTUVGDIILGLALQNLVVLQMPYGSGEQNAALLASDITYLDYLKSTE QLTEEVQSKAFFLLSNGYQRQLSFKNSDGSYSVFWQQSQKSIWLSALTFKTLERMKK YVFIDENVQKQTLIWLSSQKTSKGFKNQDGLFNHAWQGGDEEDISLTAYVVGMPFEA GAALDSGVTNGYNHAILAYAFALAGKEKQVESLLQTLQDQAPKLNVIYWERERKPKT EEFSPFIPWAPSQTEKSCYVLLAVISRKIPDLTYASKIVQWLAQRNMSSHGGFSSNQD QNTVTFSSESSEIFQVNGHNRLLVQRSEVTQAPGEYTVDVEGHGCTFIQATLKYNVL LPKKASGFSLSLIEIVKNYSSTAFDLTVTLKYTGIRNKSSMVVIDVKMLSGFTPTMSSI EELNKGQVMKTEVKNDHVLFYLENVGFGRADSFPPFSVEQSNLVFNIQAPAMVYDYY EKEEYALAFYNIDSSSVSE		
	SEQ ID NO: 125	4501 bp	
NOV26b, CG59584-02 DNA Sequence	TCCATTTCTATAGACATTAACCTGAATTAGCGCCCTCAGTAGATGCTTGTCTATA GCTTGATCCTGGAGGAGAAATGGTCACTGATAGCACCCAAATTCGAATTGAGAAATG CTTCGAAATCAGGTCAACTTAAATTTTCTAAAGAAAAAGTTTACCAGGATCCAAT ATTGATCTTCAAGTCTCGGCTGCTTCAAACCTCTCTTTGTGCTCTTTGGGCTGTAGACC AGAGTGTATTGCTACTAAGGAATTATGGTCAGCTGTGAGCACAACCTGTGTATAGTCA GCTATATTCCAGGGAACATACATGGCTATTACTTCAGAGGACTTAACCTAGAAGATGGC CTTAAAGTGCCGTGTCTTGAAGATGAACATATCCTTTACAATGGAATTTATTACACAC CTGCATGGGCTGACTTTGGAAAAGATGGCTATGACCTTGTGAAGGATCCTCAAAACAA TCGGATTTTCAAAGGCAAAATGTGACTTCTTCCGAAATATTACCAACTCTCGTTC CAACTGATTTCAAGAACCAATGTTTGGAGATTACTGGATTGTTGTGAAAGAAACTCAA GGGAGACAGTGACACACCAATTTGCTGTTAAAAGATATGTGCTGCCCAAGTTTGAAGT TACAGTCAATGCACCACAAACAGTAACTATTTTCAAGATGATGAATTTCAAGTGGATGTA TGTGCTAAGTACAACTTTGGCCAACCTGTGCAAGGGGAAACCAAACTCCGGGTGTGCA GAGAGTATTTTTCTTCAAGCAATTTGTGAGAAAAATGAAAATGAAATATGTGAGCAATT TATTGCACAGTTGGAATGGTTGTGTTTCTCAAATTGTAATACAAAGTCTTCCAA		

CTCTACCGTTCGGGATTGTTTCATGACATTTTCATGTCGCTGTAATTGTTACAGAATCTG  
GGACAGTTATGCAGATCAGCGAGAAGACCTCAGTTTTTATCACTCAATTGCTTGGAAAC  
TGTAACCTTTGAGAACATGGATACATTCTATAGAAGAGGGATTCTTATTTTGGAACT  
CTTAAATTTTCGGATCCCAATAATGTACCTATGGTGAACAAGTTGTTGCAACTGGAGC  
TCAATGATGAATTTATAGGAAATTACACTACGGATGAGAATGGCGAAGCTCAATTTTC  
CATTGACACTTCAGACATATTTGATCCAGAGTTCAACCTAAAAGCCACATATGTTTGA  
CCTGAGAGCTGCTATCTTCCAGCTGGTTGACGCCCTCAGTACTTGGATGCTCACTTCT  
TAGTCTCAGCTTTTACTCCCGAACCAACAGCTTCTGAAGATTGTTCCAGAACCAAA  
GCAGCTTGAATGTAATCAACAGAAGGTGTTACTGTGCATTACTCCCTAAACAGTGAA  
GCATATGAGGATGATTCCAATGTAAAGTTCTTCTATTTGATGATGGTAAAAGGAGCTA  
TCTTACTCAGTGGACAAAAGGAAATCAGAAACAAAGCCTGGAATGGAACTTCTCGTT  
CCCAATCAGCATCAGTGTGATCTGGCTCCTGCAGCCGCTCCTGTTGCTATACCCCTT  
CACCCAGTGGGGAATTTGGCTGACAGTGTGAGATTCCAGGTTGACAAGTGCTTTA  
AACACAAGGTTAACATAAAGTTCTCTAACGAGCAGGGCTTACCTGGTTCCAATGCTAG  
TCTCTGTCTTCAAGCGCGCTGTCTTATTCTGTGCCCTCAGGGCTGTGGATAGGAAT  
GTCCTTCTACTGAAATCTGAACAACAGCTGTGAGCTGAAAGTGTGTATAACATGGTTC  
CAAGTATAGAGCCGTATGGTTATTTCTACCATGGCCTCAATCTTGTATGGCAAGGA  
AGACCCTTGCAATCCTCAGAGGGATATGTTCTACAATGGTTTATATTACACACCTGTA  
AGCAACTATGGGGATGGAGATATCTATAATATTGTGAGGAACATGGGTCTAAAAGTCT  
TTACCAATCTCCATTACCGAAAACCAGAAAAAATATGGTCCAATGCGTAGTGTCCG  
TCTAGAATTGCATGTAGCTAGTGGAAATCAGAGGGGAGAATGCTGACTATGTAGAACAG  
GCTATAATTCAAACAGTAAGAACAACCTTCCAGAGACATGGATGTGGGACCTCGTCA  
GTGTGATTCCTCAGGCTCTGCCAATCTTTCGTTCTCATTCTGATACGATAACCCA  
ATGGGAGGCAAGTGGCTTTTGTGTGAATGGTGACGTTGGATTGGCATTTCCTCTACA  
ACCACTCTAGAAGTCTCCCAACCTTTCTTTATTGAGATTGCCTCACCTTTTCGGTTG  
TTCAAAATGAACAATTTGATTGATTGTCAATGTCTTCAGCTACCGGAATACATGTGT  
AGAGATTTCTGTTCAAGTGGAGGAGTCTCAGAATTATGAAGCAAATATTATACCTTG  
AAAATCAATGGCAGTGAGGTTATTTCAAGCTGGAGGGAGGAAAAACAACGCTCTGGACTA  
TTATACCTAAGAAATTGGGTAAAGTGAATATCACTGTAGTTGCTGAGTCCAAACAAAG  
CAGTGTCTGCCCCAATGAAGGAATGGAGCAGCAAAAGCTAAACTGGAAGACACTGTG  
GTCCAAAGCTTCTTAGTAGAGCCTGAAGGTATTGAAAAGGAAAGGACCAGAGTTCC  
TTATCTGTACAGAAGGTGCCAAAGCCTCCAAGCAGGGAGTTTGGACTTGCCAAACGA  
TGTAAGTAAGGGTCAGCCAGAGGCTTTTCACTGTTGTGGGGATATTCTAGGACTT  
GCCTTGCAAGATCTGGTTGTTCTCCAATGCCCTATGGAAGTGGAGAGCAGAATGCTG  
CCCTACTAGCATCTGATACTTATGTTCTGGACTATCTGAAATCTACTGAGCAACTGAC  
AGAGGAAGTTCAATCTAAGGCTTTCTTCTCTTATCTAATGTTTATCAAGGCAATTA  
TCTTTCAAAACTCTGATGGTTCTATAGTGTGTTTGGCAGCAGAGTCAGAAAGGAA  
GCATATGGCTCAGTGTCTTACTTTTAAGACATTGGAGAGAATGAAAAATATGTATT  
CATTGATGAAAAATGTTCAAAAACAGACCTTAATCTGGCTTTCAAGCCACACAGAAAA  
AGCGGCTGCTTTAAGAATGATGGCCAGCTTTTCAACCACGCCTGGGAGGGTGGAGATG  
AAGAGGACATTTCACTCACTGCGTATGTTGTTGGGATGTTCTTTGAAGCTGGGCTCAA  
TTTCACTTTTCCGTCTACGAAACGCACTCTTTGCCTTGAAGCGGCAATTGGACAGT  
GGTGTCACTAATGGCTATAATCATGCAATTCAGCTTATGCTTTTGCCTTAGCTGGAA  
AAGAGAAGCAAGTGAATCTTTACTCCAACCCCTGGATCAATCTGCCCAAAACTAAA  
TAATGTCATCTACTGGGAAAGAGAAAGGAAACCAAGACAGAAGAAATTTCCATCCTTT  
ATTCCCTGGGCACCTTCTGCTCAGACTGAGAAGAGTTGCTACGTGCTGTTGGCTGTCA  
TTTCCCGGAAAAATCCTGACCTCACCTATGCTAGTAAGATTGTGACGTGGCTTGCCCA  
ACGGATGAATTTCCCATGGAGGCTTTTCTTCCAACCAGACACCTGATGATACTCTGTTT  
AAATTATATACGGGCCAAAAAGAAAGCTTTGCTCTAGTTCTGTGGGCTATACACTGG  
GAAAAGCAAATGAAAAGAAAGGAAAACAGGAGAAATGGGGTGAAGGATCCAGTGAGAT  
TTTCCAGGTTAACGGTCATAACCGCCTACTGGTCCAACGTTCAGAAGTAACACAGGCA  
CCTGGAGAATACACAGTAGATGTGGAAGGACACGGTTGTACATTTATCCAGGCCACCC  
TTAAGTACAATGTTCTCTACCTAAGAAGGCATCTGGATTCTCTTCTTCTTGGAAAT  
AGTAAAGAACTACTCTCGACTGCTTTTGACCTCAGTGACCCTCAAATACACTGGA  
ATTGCAATAAAATCCAGTATGGTGGTTATAGATGTAATAATGCTATCAGGATTTACTC  
CAACCATGTATCCATTGAAGAGCTTGAAAACAAGGGCCAAGTGATGAAGACTGAAGT  
CAAGAATGACCATGTTCTTTCTACTTGGAAAATGTAGGTTTGGTTCGAGCAGACAGT  
TTCCCTTTTCTGTTGAGCAGAGCAACCTTGTGTTCAACATTCAGCCAGCCCCAGCCA  
TGGTCTACGATTATTATGAAAAGAAAGAAATATGCCCTAGCTTTTACAACATCGACAG  
TAGTTCAGTTTCCGAGTGAGACAAAGCAATTACTAGAAGAGTTGGAGAAGCATTCTT  
GTAACAACTGATTCTTCTGTATCAAACCTGGAAAAAATCATGAACCATCTGACATC

	GTGAACAGTCTGCAGTGGGTATGGTTCTTGTCAAGTCTTATTCCTTATCATCCCA TTAAATGTTGTCATTTTGCAAAAAAAAAAAAAA		
	ORF Start: TCC at 1	ORF Stop: TGA at 4309	
	SEQ ID NO: 126	1436 aa	MW at 161836.4kD
NOV26b, CG59584-02 Protein Sequence	SISIDINPELAPSVDMLVYSLHPGGEMVTDSTQFRIEKCENQVNLNFSKEKSLPGSN IDLQVSAASNSLCALWAVDQSVLLLRNYGQLSAQTVYSQLYSRELHGYYFRGLNLEDG LKVPCLEDEHILYNGIYYTPAWADFGKDGVDLVKDPQNNRI FQRQNVTSFRNITQLSF QLISEPMFGDYWIVVKRNSRETVTTHQFAVKRYVLPKFVTVNAPQTVTISDDEFQVDV CAKYNFGQPVQGETQIRVCREYFSSSNCEKNENEICEQFIAQLENGCVSQIVNTKVFQ LYRSGLFMTFHVAVIVTESGTVMQISEKTSVFITQLLGTNVFENMDTFYRRGISYFGT LKFSDPNNVPMVNKLLQLELNDEFIGNYTTDENGAEQFSIDTSDIFDPEFNLKATYVR PESCYLPSWLTPOYLDAHFLVSRFYSRTNSFLKIVPEPKQLECNQKQVTVVHYSLNSE AYEDDSNVKFFFLMMVKGAILLSGQKEIRNKAWNGNFSFPISISADLAPAAVLVFTYL HPSGEIVADSVRFQVDKCFKHVNIKFSNEQGLPGSNASLCLQAAPVLFALRAVDRN VLLKSEQQLSAESVYNMVPSIEPYGYFYHGLNLDGKEDPCIPQRDMFYNGLYTPV SNYGDGDIYNI VRNMGLKVFTNLHYRKPEKIMVQC VVFRLELHVASGIRGENADYVEQ AIIQTVRTNFPETWMWDLVSVDSGSSANLSFLIPDTITQWEASGFCVNGDVGFISST TTLEVSQPFIEIASPFSVVQNEQFDLIVNVFSYRNTCVEISVQVEESQNYEANIHTL KINGSEVIQAGGRKTNVWTIIPKKLGKVNITVVAESKQSSACPNEGMEQQKLNWKDTV VQSFLVEPEGIEKERTQSFLICTEGAKASKQGVLDLPNDVVEGSARGFTTVVGDILGL ALQNLVVLQMPYGSGEQNAALLASDTYVLDYKSTEQLTEEVQSKAFFLLSNGYQRQL SFKNSDGSYSVFWQSQKSIWLSALTFTKTLERMKKYVFIDENVQKQTLIWLSSQKKT SGCFKNDGQLFNHAWEGGDEEDISLTAYVVGMMFEAGLNFTFPALRNALFCLEALDS GVTNGYNHAILAYAFALAGKEKQVESLLQTLDSAPKLNNVIYWERERKPKTEFPSPF IPWAPSAQTEKSCYVLLAVISRKI PDLTYASKIVQWLAQRMNSHGGFSSNQTPDDTLF KLYTGQKESFRSSSVGYTLGKANEEKENRRNGGEGSSEIFQVNGHNRLLVQRSEVTQA PGEYTVDVEGHGCTFIQATLKYNVLLPKKASGFSLSLEIVKNYSSTAFDLTVTLKYTG IRNKSSMVVIDVKMLSGFTPTMSSI EELNKGQVMKTEVKNDHVLFYLENVGFGGRADS FFFSVEQSNLVFNIQPAPAMVYDYYEKEEYALAFYNIDSSSVSE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 26B.

Table 26B. Comparison of NOV26a against NOV26b and NOV26c.		
Protein Sequence	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV26b	164..1411 150..1436	997/1311 (76%) 1072/1311 (81%)

Further analysis of the NOV26a protein yielded the following properties shown in Table 26C.

Table 26C. Protein Sequence Properties NOV26a	
PSort analysis:	0.8200 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1380 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 46 and 47

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 26D.

<b>Table 26D. Geneseq Results for NOV26a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV26a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAB50673	Human alpha-2 macroglobulin protein SEQ ID NO:59 - Homo sapiens, 1474 aa. [WO200073328-A2, 07-DEC-2000]	35..1407 30..1468	552/1458 (37%) 846/1458 (57%)	0.0
AAY97157	Human alpha-2-macroglobulin - Homo sapiens, 1474 aa. [WO200046246-A1, 10-AUG-2000]	35..1407 30..1468	551/1458 (37%) 846/1458 (57%)	0.0
AAR11334	Recombinant human alpha-2 macroglobulin - Homo sapiens, 1474 aa. [WO9103557-A, 21-MAR-1991]	35..1407 30..1468	549/1458 (37%) 844/1458 (57%)	0.0
AAR11749	Human alpha-2 macroglobulin bait region mutant - Homo sapiens, 1484 aa. [WO9103557-A, 21-MAR-1991]	35..1407 30..1478	546/1460 (37%) 842/1460 (57%)	0.0
AAB43949	Human cancer associated protein sequence SEQ ID NO:1394 - Homo sapiens, 1285 aa. [WO200055350-A1, 21-SEP-2000]	187..1407 2..1279	497/1295 (38%) 753/1295 (57%)	0.0

- In a BLAST search of public sequence databases, the NOV26a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 26E.

<b>Table 26E. Public BLASTP Results for NOV26a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV26a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P20740	Ovostatin precursor (Ovomacroglobulin) - Gallus gallus (Chicken), 1473 aa.	1..1402 1..1461	640/1482 (43%) 931/1482 (62%)	0.0
P01023	Alpha-2-macroglobulin precursor (Alpha-2-M) - Homo sapiens (Human), 1474 aa.	35..1407 30..1468	552/1458 (37%) 846/1458 (57%)	0.0
CAA01532	ALPHA 2-MACROGLOBULIN	35..1407 30..1468	550/1458 (37%) 845/1458 (57%)	0.0

	1474 aa.			
P06238	Alpha-2-macroglobulin precursor (Alpha-2-M) - Rattus norvegicus (Rat), 1472 aa.	26..1408 13..1467	552/1477 (37%) 852/1477 (57%)	0.0
CAA01533	ALPHA 2-MACROGLOBULIN 690-740 - Homo sapiens (Human), 1484 aa.	35..1407 30..1478	547/1460 (37%) 844/1460 (57%)	0.0

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26F.

Table 26F. Domain Analysis of NOV26a			
Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
A2M_N: domain 1 of 1	35..611	178/655 (27%) 381/655 (58%)	3.3e-96
A2M: domain 1 of 3	717..1096	137/414 (33%) 268/414 (65%)	2.2e-95
prenyltrans: domain 1 of 1	1194..1214	7/21 (33%) 15/21 (71%)	4.4
A2M: domain 2 of 3	1114..1218	45/110 (41%) 72/110 (65%)	1e-19
A2M: domain 3 of 3	1226..1402	61/242 (25%) 125/242 (52%)	1.1e-35

#### EXAMPLE 27.

The NOV27 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis		
	SEQ ID NO: 127	880 bp
NOV27a, CG59417-01 DNA Sequence	ACTCACTATAGGGCTCGAGCGGCACCATGGCTTTCCTCTGGCTCCTCTCCTGCTGGGC CCTCCTGGGTACCACTTCGGCTGCGGGTCCCCGCCATCCACCCTGTGTTACGCGGC CTGTCCAGGATCGTGAATGGGAGGACGCCGTCCCGGCTCCTGGCCCTGGCAGGTGT CCCTGCAGGACAAAACCGCTTCCACTTCTGCGGGGCTCCCTCATAGCGAGGACTG GGTGGTCACCGCTGCCACTGCGGGTCAAGACCTCCGACGTGGTCTGGCTGGGGAG TTTGACCAGGGCTCTGACGAGGAGAACATCCAGGTCTGAAGATCGCCAAGTCTTCA AGAACCCCAAGTTCAGCATTCGACCGTGAACAATGACATCACCTGCTGAAGCTGGC CACACCTGCCGCTTCTCCAGACAGTGTCCGCCGTGTGCCTGCCAGCGCCGACGAC GACTTCCCCGCGGGACACTGTGTGCCACCACAGGCTGGGGCAAGACCAAGTACAACG CCAACAAGACCCCTGACAAGCTGCAGCAGGCAGCCCTGCCCTCCTGTCCAATGCCGA ATGCAAGAAGTCTGGGGCAGGAGGATCACCGACGTGATGATCTGTGCCGGGGCCAGT GGCGTCTCCTCCTGCATGGGTGACTCTGGAGGCCCTGGTCTGCCAGAAGGACGGAG CCTGGACCCTGGTGGGCATTGTGTCTGGGGCAGCCGCACCTACTCTACCACCACGCC	

	CGCTGTGTACGCCCGTGTACCAAGCTCATACCCTGGGTGCAGAAGATCCTGGCCGCC AACTGAGCCCGCAGCTCCTGCCACCCTGCCTTAAGATTCCCATTAAATGCATCTGT TTAGAAAAAA		
	ORF Start: ATG at 27	ORF Stop: TGA at 816	
	SEQ ID NO: 128	263 aa	MW at 28046.9kD
NOV27a, CG59417-01 Protein Sequence	MAFLWLLSCWALLGTTFGCGVPAIHPVFSGLSRIVNGEDAVPGSWPQVSLQDKTGFH FCGGSLSISEDWVVTAAHCGVRTSDVVVAGEFDQGSDEENIQVLKIAKVFNPKFSILT VNNDITLLKLATPARFSQTVSAVCLPSADDDFPAGTLCATTGWGKTKYNANKTPDKLQ QAALPLLSNAECKKSWGRRITDVMICAGASGVSSCMGDSGGPLVCQKDGAWTLVGIVS WGSRTYSTTTPAVYARVTKLIPWVQKILAA		

Further analysis of the NOV27a protein yielded the following properties shown in Table 27B.

Table 27B. Protein Sequence Properties NOV27a	
PSort analysis:	0.3700 probability located in outside; 0.1040 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 19 and 20

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

- 5 homologous proteins shown in Table 27C.

Table 27C. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB98504	Human chymotrypsin serine protease catalytic domain - Homo sapiens, 231 aa. [WO200129056-A1, 26-APR-2001]	33..263 1..231	226/231 (97%) 228/231 (97%)	e-132
AAV99596	Bovine chymotrypsinogen A - Bos taurus, 245 aa. [WO200032759-A1, 08-JUN-2000]	19..263 1..245	197/245 (80%) 213/245 (86%)	e-116
AAB11711	Mouse serine protease BSSP5 (mBSSP5) SEQ ID NO:4 - Mus sp, 264 aa. [WO200031243-A1, 02-JUN-2000]	1..263 1..264	150/264 (56%) 188/264 (70%)	5e-87
AAB11710	Human serine protease BSSP5 (hBSSP5) SEQ ID NO:2 - Homo sapiens, 264 aa. [WO200031243-A1, 02-JUN-2000]	1..263 1..264	141/264 (53%) 184/264 (69%)	2e-82

AAB54190	Human pancreatic cancer antigen protein sequence SEQ ID NO:642 - Homo sapiens, 133 aa. [WO200055320-A1, 21-SEP-2000]	132..263 2..133	127/132 (96%) 129/132 (97%)	1e-71
----------	---	--------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27D.

Table 27D. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P17538	Chymotrypsinogen B precursor (EC 3.4.21.1) - Homo sapiens (Human), 263 aa.	1..263 1..263	257/263 (97%) 259/263 (97%)	e-152
P04813	Chymotrypsinogen 2 precursor (EC 3.4.21.1) - Canis familiaris (Dog), 263 aa.	1..263 1..263	228/263 (86%) 241/263 (90%)	e-135
Q9CR35	2200008D09RIK PROTEIN - Mus musculus (Mouse), 263 aa.	1..263 1..263	223/263 (84%) 246/263 (92%)	e-135
P07338	Chymotrypsinogen B precursor (EC 3.4.21.1) - Rattus norvegicus (Rat), 263 aa.	1..263 1..263	222/263 (84%) 244/263 (92%)	e-135
Q9DC86	2200008D09RIK PROTEIN - Mus musculus (Mouse), 263 aa.	1..263 1..263	222/263 (84%) 246/263 (93%)	e-134

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27E.

Table 27E. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value
trypsin: domain 1 of 1	34..256	109/261 (42%) 194/261 (74%)	5.6e-102

5

#### EXAMPLE 28.

The NOV28 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis		
	SEQ ID NO: 129	1749 bp



NOV28a, CG59415-01 DNA Sequence	GCGGTCCCCAGCCTGGGTAAGATGGCCCCATGGCCCCGAAGGGCCTAGTCCCAGCT GTGCTCTGGGGCCTCAGCCTCTTCTCAACCTCCCAGGACCTATCTGGCTCCAGCCCT CTCCACCTCCCCAGTCTTCTCCCCGCCTCAGCCCCATCCGTGTCTATACCTGCCGGG ACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGACAACCTTTGGAGGT GGAAACACTGCCTGGGAGGAAGAGAAATTTGTCCAAATACAAAGACAGTGAGACCCGCC TGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGTGCCACCGCCTGT GGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTCAACAAGCAGCAGGAGGCCCG GACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGCCCCGAGGCACCT TCGGGCCCTCCTGCCTTCCCTGTCTGGGGGAACAGAGAGGCCCTGCGGTGGCTACGG GCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCAGTGTGACTGCCAAGCCGGC TACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAGGCAGAACGCAACG CCAGCCATCTGGTATGTTCCGGCTTGTTTTGGCCCTGTGCCGATGCTCAGGACCTGA GGAATCAAACCTGTTTGAATGCAAGAAGGGCTGGGCCCTGCATCACCTCAAGTGTGTA GACATTGATGAGTGTGGCACAGAGGGAGCCAACTGTGGAGCTGACCAATTCTGCGTGA ACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCCTAGGCTGCATGGG GGCAGGGCCAGGTGCTGTGAAGAAGTGTAGCCCTGGCTATCAGCAGGTGGGCTCCAAG TGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCCGGAGAGAAACAAGCCAGT GTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACAAGCAGATGGA AGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTCAGAGATGACA GAAGACGAGTTGGTGGTGTGTCAGCAGATGTTCTTTGGCATCATCATCTGTGCACTGG CCACGCTGGCTGCTAAGGGGGACTTGGTGTTCACCGCCATCTTCATTGGGGCTGTGGC GGCCATGACTGGGTACTGGTGTGTCAGAGCGCAGTGACCGTGTGCTGGAGGGCTTCATC AAGGGCAGATAATCGCGGCCACCACCTGTAGGACCTCCTCCACCCAGCTGCCGCCCA GAGCTTGGGCTGCCCTCCTGCTGGACACTCAGGACAGCTTGGTTTATTTTGAAGTG GGGTAAACACCCCTACCTGCCTTACAGAGCAGCCAGGTACCCAGGCCCGGGCAGACA AGGCCCTGGGGTAAAAAGTAGCCCTGAAGGTGGATACCATGAGCTCTTACCTGGCG GGGACTGGCAGGCTTACAATGTGTGAATTTCAAAGTTTCTTAAATGGTGGCTGC TAGAGCTTTGGCCCTGCTTAGGATTAGGTGGTCTCACAGGGGTGGGGCCATCACAG CTCCCTCCTGCCAGCTGCATGCTGCCAGTTCCTGTTCTGTGTTACCATCCCCACA CCCCATTGCCACTTATTTATTCATCTCAGGAAATAAGAAAGGTCTTGGAAAGTTAAA AAAAAAA
	ORF Start: ATG at 23      ORF Stop: TAA at 1286
	SEQ ID NO: 130      421 aa      MW at 45520.1kD
NOV28a, CG59415-01 Protein Sequence	MAPWPPKGLVPAVLWGLSLFLNLPPIWLQSPFPQSSPPQPHPCHTCRGLVDSFNK GLERTIRDNFNGGNTAWEEENLSKYKDSERLVEVLEGVCSKSDFECHRLLELSEELV ESWWFHKQOEAPDLFQWLCSDSLKLCPPAGTFGPSCLPCPGQTERPCAGYGQCEGEGT RGGSGHCDQCQAGYGGEACGQCGLGYPEAERNASHLVCSACFGPCARCSGPEESNCLQC KKGWALHHLKCVDIDECETEGANCGADQFCVNTGEGSYECRDCAKACLCMGAGPGRCK KCSPGYQQVGSKCLDVDECETEVCPGREQAQCENTEGGYRCICAEGYQMEGICVKBO IPESAGFFSEMTEDELVLVQMFPGIIICALATLAAKGDVLFTAIFIGAVAAMTGYWL SERSDRVLEGFIKGR
	SEQ ID NO: 131      1011 bp
NOV28b, 191815704 DNA Sequence	GGATCCCAGCCCTCTCCACCTCCCCAGTCTTCTCCCCGCCTCAGCCCCATCCGTGTC ATACCTGCCGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGA CAACTTTGGAGGTGGAACACTGCCTGGGAGGAAGAGAAATTTGTCCAAATACAAAGAC AGTGAGACCCGCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGT GCCACCGCCTGCTGGAGCTGAGTGAGGAGCTGGTGGAGAGCTGGTGGTTTCAAGCA GCAGGAGGCCCCGACCTCTTCCAGTGGCTGTGCTCAGATTCCTGAAGCTCTGTGTC CCCGCAGGCACCTTCGGGGCCTCCTGCCTTCCCTGTCTGGGGGAACAGAGAGGCCCT GCGGTGGCTACGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCAGTGTGA CTGCCAAGCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGCTACTTTGAG GCAGAACGCAACGCCAGCCATCTGGTATGTTCCGGCTTGTTTTGGCCCTGTGCCGAT GCTCAGGACCTGAGGAATCAAACCTGTTTGAATGCAAGAAGGGCTGGGCCCTGCATCA CCTCAAGTGTGTAGACATTGATGAGTGTGGCACAGAGGGAGCCAACCTGTGGAGCTGAC CAATTCTGCGTGAACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCC TAGGCTGCATGGGGCAGGGCCAGGTGCTGTGAAGAAGTGAAGCCCTGGCTATCAGCA GGTGGGCTCCAAGTGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAG AACAAGCAGTGTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACA

	AGCAGATGGAAGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTC AGAGATGACAGAAGACGAGCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 132	337 aa	MW at 36352.1kD
NOV28b, 191815704 Protein Sequence	GSQPSPPPPQSSPPPPQHPCHTCRGLVDSFNKGLERTIRDNFGGGNTAWEEENLSKYKD SETRLVEVLEGVCSKSDFECHRLLELSEELVESWWFHKQQEAPDLFQWLCSDSLKLC PAGTFGPSCLPCPGGTERPCGGYGQCEGEGTRGSGHCDQCAGYGGAEACQCGLGYFE AERNASHLVCSACFGPCARCSGPESNCLQCKKGWALHHLKCVDIDECGTGEGANCGAD QFCVNTEGSYECRDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLDVECETEVCPGE NKQCENTEGGYRCICAEGYKQMEGICVKEQIPESAGFFSEMTEDELE		
	SEQ ID NO: 133	1011 bp	
NOV28c, 191815724 DNA Sequence	GGATCCCAGCCCTCTCCACCTCCCAAGTCTTCTCCCCCGCCTCAGCCCCATCCGTGTC ATACCTGCCGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGA CAACTTTGGAGGTGAAACACTGCCTGGGAGGAAGAGAATTTGTCCAAATACAAAGAC AGTGAGACCCGCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGT GCCACCGCTGCTGGAGCTGAGTGAGGAGCTGGTGAGAGCTGGTGGTTTCAACAAGCA GCAGGAGGCCCGGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGC CCCGCAGGCACCTTCGGGCCCTCCTGCCTTCCCTGTCTCTGGGGGAACAGAGAGGCCCT GCGGTGGCTGCGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCACTGTGA CTGCCAAGCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAG GCAGAACGCAACGCCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCGTGTGCCCGAT GCTCAGGACCTGAGGAATCAAACCTGTTTGCAATGCAAGAAGGGCTGGGCCCTGCATCA CCTCAAGTGTGTAGACATTGATGAGTGTGGCACAGAGGGAGCCAAGTGTGGAGCTGAC CAATTCTGCGTGAACTGAGGGCTCCTATGAGTGCAGAGACTGTGCCAAGGCCTGCC TAGGCTGCATGGGGGCAGGGCCAGGTCGCTGTAAGAAGTGTAGCCCTGGCTATCAGCA GGTGGGCTCCAAGTGTCTCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAG AACAAGCAGTGTGAAAACACCGAGGGCGGTTATCGCTGCATCTGTGCCAGGGCTACA AGCAGATGGAAGGCATCTGTGTGAAGGAGCAGATCCCAGAGTCAGCAGGCTTCTTCTC AGAGATGACAGAAGACGAGCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 134	337 aa	MW at 36292.1kD
NOV28c, 191815724 Protein Sequence	GSQPSPPPKSSPPPPQHPCHTCRGLVDSFNKGLERTIRDNFGGGNTAWEEENLSKYKD SETRLVEVLEGVCSKSDFECHRLLELSEELVESWWFHKQQEAPDLFQWLCSDSLKLC PAGTFGPSCLPCPGGTERPCGGCGQCEGEGTRGSGHCDQCAGYGGAEACQCGLGYFE AERNASHLVCSACFGPCARCSGPESNCLQCKKGWALHHLKCVDIDECGTGEGANCGAD QFCVNTEGSYECRDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLDVECETEVCPGE NKQCENTEGGYRCICAEGYKQMEGICVKEQIPESAGFFSEMTEDELE		
	SEQ ID NO: 135	1646 bp	
NOV28d, CG59415-02 DNA Sequence	GGCGACGCGGTCCCCAGCCTGGGTAAAGATGGCCCCATGGCCCCGAAGGGCCTAGTC CCAGCTGTGCTCTGGGGCCTCAGCCTCTTCTCAACCTCCCAGGACCTATCTGGCTCC AGCCCTCTCCACCTCCCCAGTCTTCTCCCCCGCCTCAGCCCCATCCGTGTACATCTG CCGGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGACAACCTT GGAGGTGGAACACTGCCTGGGAGGAAGAGAATTTGTCCAAATACAAAGACAGTGAGA CCCGCCTGGTAGAGGTGCTGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGTGCCACCG CCTGCTGGAGCTGAGTGAGGAGCTGGTGAGAGCTGGTGGTTTCAACAAGCAGCAGGAG GCCCCGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAGCTCTGCTGCCCGCAG GCACCTTCGGGGCCTCCTGCCTTCCCTGTCTCTGGGGGAACAGAGAGGCCCTGCGGTGG CTACGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCACTGTGACTGCCAA GCCGGCTACGGGGGTGAGGCCTGTGGCCAGTGTGGCCCTTGGCTACTTTGAGGCAGAAC GCAACGCCAGCCATCTGGTATGTTTCGGCTTGTTTTGGCCCCGTGTGCCCGATGCTCAGG ACCTGAGGAATCAAACCTGTTTGCAATGCAAGAAGGGCTGGGCCCTGCATCACCTCAAG TGTGTAGACTGTGCCAAGGCCTGCCTAGGCTGCATGGGGGCAGGGCCAGGTCGCTGTA AGAAGTGTAGCCCTGGCTATCAGCAGGTGGGCTCCAAGTGTCTCGTGAGTCTCTGCT GATGGACACAGGCACCGGCTCACCCAGCATGAATGGTGAAGAGGCTGGAATATGGGCA GGTGGGGGAAGGAAGGGTGAATGTTGCCTGGGCAGAGGGGAGGAGATGGACAAGATG		

	GAGTCAGGTGCTGGGTGGGGGGCCCTAGCAGGACTCTGACCCCTCCCTCCCCTCAAGA TGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAGAACAAGCAGTGTGAAAACACC GAGGGCGGTTATCGCTGCATCTGTGCCGAGGGCTACAAGCAGATGGAAGGCATCTGTG TGAACAGAAGACGAGTTGGTGGTGTGCAGCAGATGTTCTTTGGCATCATCATCTGTG CACTGGCCACGCTGGCTGCTAAGGGCGACTTGGTGTTCACCGCCATCTTCATTGGGGC TGTGGCGGCCATGACTGGCTACTGGTTGT CAGAGCGCAGTGACCGTGTGCTGGAGGGC TTCATCAAGGGCAGATAATCGCGGCCACCACCTGTAGGACCTCCTCCACCCACGCTG CCCCCAGAGCTTGGGCTGCCCTCCTGCTGGACACTCAGGACAGCTTGGTTTATTTTG AGAGTGGGGTAAGCACCCCTACCTGCCTTACAGAGCAGCCAGGTACCCAGGCCGGG CAGACAAGGCCCCCTGGGGTAAAAAGTAGCCCTGAAGGTGGATACCATGAGCTCTTCAC CTGGCGGGGACTGGCAGGCTTCACAATGTGTGAATTCAAAGTTTTTCTTAATGGTG GCTGCTAGAGCTTTGGCCCCCTG		
	ORF Start: ATG at 29	ORF Stop: TAA at 1238	
	SEQ ID NO: 136	403 aa	MW at 42961.2kD
NOV28d, CG59415-02 Protein Sequence	MAPWPPKGLVPAVLWGLSLFLNLP GPIWLQPSPPPPQSPPPQPHPCHTCRGLVDSFNK GLERTIRDNFGGNTAWEENLSKYKDSERLVEVLEGVCSKSDFECHRLLELSEELV ESWWFHKQQEAPDLFQWLCSDSLKLCCPAGTFGPSCLPCPGTERPCGGYGQCEGEGT RGGSGHCDQCAGYGGEACGQCGLGYFEAERNASHLVCSACFGPCARCSGPESNCLQC KKGWALHHLKCVDCAKACLGCMGAGPGRCKKCSPGYQQVGSKCLVSLLLMDTGTGSPS MNGEEAGIWAGGGRKGMLPGQRGGDGDGVRCWVGGPSRTLTPSPQDVDECETEVC PGENKQCENTEGGYRCICAEGYKQMEGICVNRVRVGGAAADVLWHHHLCTGHAGC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

Table 28B. Comparison of NOV28a against NOV28b through NOV28d.		
Protein Sequence	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV28b	44..364 17..336	307/321 (95%) 307/321 (95%)
NOV28c	46..364 19..336	286/319 (89%) 286/319 (89%)
NOV28d	1..348 1..381	270/384 (70%) 275/384 (71%)

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 30 and 31

- 5 A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 28D.

<b>Table 28D. Geneseq Results for NOV28a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV28a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAU12316	Human PRO214 polypeptide sequence - Homo sapiens, 420 aa. [WO200140466-A2, 07-JUN-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAM41685	Human polypeptide SEQ ID NO 6616 - Homo sapiens, 513 aa. [WO200153312-A1, 26-JUL-2001]	1..421 94..513	418/421 (99%) 418/421 (99%)	0.0
AAM39899	Human polypeptide SEQ ID NO 3044 - Homo sapiens, 420 aa. [WO200153312-A1, 26-JUL-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAB68594	PRO214 - Homo sapiens, 420 aa. [WO200105836-A1, 25-JAN-2001]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0
AAB27228	Human EXMAD-6 SEQ ID NO: 6 - Homo sapiens, 420 aa. [WO200068380-A2, 16-NOV-2000]	1..421 1..420	418/421 (99%) 418/421 (99%)	0.0

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

<b>Table 28E. Public BLASTP Results for NOV28a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV28a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
Q9Y409	HYPOTHETICAL 44.9 KDA PROTEIN - Homo sapiens (Human), 417 aa.	1..418 1..417	413/418 (98%) 413/418 (98%)	0.0
Q91XD7	UNKNOWN (PROTEIN FOR MGC:18896) - Mus musculus (Mouse), 420 aa.	1..421 1..420	383/421 (90%) 405/421 (95%)	0.0
Q96HD1	UNKNOWN (PROTEIN FOR MGC:8447) - Homo sapiens (Human), 422 aa.	1..362 1..361	348/362 (96%) 353/362 (97%)	0.0
Q9CYA0	5730592L21RIK PROTEIN - Mus musculus (Mouse), 350 aa.	33..346 16..330	154/316 (48%) 200/316 (62%)	e-100
Q60438	HT PROTEIN - Cricetus griseus (Chinese hamster), 348 aa.	9..339 3..324	156/333 (46%) 202/333 (59%)	4e-97

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_EGF: domain 1 of 1	168..211	11/60 (18%) 32/60 (53%)	0.11
zf-MYND: domain 1 of 1	218..243	10/43 (23%) 15/43 (35%)	4.7
PHD: domain 1 of 1	217..277	12/64 (19%) 38/64 (59%)	2.8
TIL: domain 1 of 1	249..309	17/79 (22%) 37/79 (47%)	8.1
Furin-like: domain 1 of 1	189..310	36/188 (19%) 76/188 (40%)	6.5
EB: domain 1 of 1	292..344	15/62 (24%) 35/62 (56%)	0.3

**EXAMPLE 29.**

The NOV29 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis		
	SEQ ID NO: 137	6997 bp
NOV29a, CG59297-01 DNA Sequence	ATGGGGTGGAGGGGGCTGATCGCAGCCTTGCCCTGCTCTCCTTGGTGCAGCCTGCTC TGGGAACCAGCTCAAAGGATGAGGATGTAGGAAGAAGCTGGTCTGCTGACTGTCATAC TTGTGACCAGCTTGCACAGGACATGGCCGAGGAGGCAGCCAGAACATTTCTGATGAC CAGGAAAGGTGTCTCCAGGCTGCCTGCTGCCTTTCCTTTGGTGGTGAGCTGTCTGTGA GCACTGACAAGAGCTGGGGTCTTCATCTGTGCAGCTGTAGCCCTCCTGGAGGTGGATT GTGGGTGAGGTCTATGCTAATCATGTGCTTCTTATGAGTGATGGGAAGTGTGGCTGT CCTTGGTGTGCTCTGAATGGAAAGGCAGAAGACCGGGAATCACAGAGCCCATCCTCAT CAGCTTCCAGGCAGAAGAACATTTGGAACAACACTAGTGAAGCAGCGTTAAGTGTGT TAATGAAAAAACACAGGCTGTTGTTAATGAAAAAACACAGGCGCCTCTGGATTGTGAT AACAGTGCTGATAGTCTCCGAGTCTTTGCTGACAGCAGTATTGGGGAGAATTGGACCC TTCAGATGGTTTGTGACCCAGACACTTGGATGCGTGGGCCAGCTCCACGGCCTTCC GCCTGGCATTCTCGCACCCCCAGCTTCACGGCATCGCAGTCTGGTCTGAGATCCTC TATCCCCCTACTCAGCATCCTCCTGTGGCCATCCTAGCTCGAAATTCTGATAACTTCA TGAACCCTGTTCTTAATTGCTCCCTGGAAGTGGAAGCTCGGGCACCTCCAAATCTGGG ATTCCGTGTTCATATGGCTTCTGGAGAGGCTCTCTGTCTGATGATGATTTCCGGGAC AGTTCTGGGGTTGAAATGAGGCTACACAACATGTCTGAGGCAATGGCGGTGACTGCCT ACCACCAGTACTCAAAGAAGGAGTCTATATGCTCAAGGCTGTTATTATAACGAGTT TCATGGAACCGAAGTGGAGCTTGGGCCTTATTATGTGGAGATTGGCCATGAGGCCGTG TCTGCGTTCATGAACTCCAGCAGTGTCCATGAAGATGAAGTGCTTGTCTTTGCTGACT CCCAAGTGAATCAGAAACCGTCTCTGTCTACACAAATGGAAGTGTGTTGGCCACAGA CACAGACATTACATTTACAGCTGTTACCAAGGAAACAATACCCCTGGAATTTGAGTGG	

TATTTTGGAGAGGATCCACCAGTGAGGACAACCTTCAAGAAGCATTAAAAAAGACTCA  
 GCATCCCCCAATGGTATCGTGTGATGGTTAAGGCTTCCAACAGGATGAGCAGTGTGGT  
 CTCTGAGCCCCATGTCATCAGGGTGCAGAAGAAAATTGTGGCCAATCGGCTCACGTCC  
 CCCTCCTCAGCTCTGGTAAATGCCAGTGTGGCCTTTGAGTGCTGGATCAACTTCGGCA  
 CAGATGTTGCCTACCTGTGGGACTTTGGGGATGGCACCCTGAGCCTGGGGAGCAGCTC  
 CAGCAGCCATGTCTACAGTAGGGAAGGAGAATTACAGTGAGGTCCTTGCCCTTCAAT  
 AATGTCAGTGCCTCCACTCTAAGACAGCAACTTTTCATCGTGTGCGAGCCCTGCCAGC  
 CACCCCTGGTGAAGAACATGGGGCCTGGGAAAGTCAGATATGGAGGTCTCAGCCTGT  
 GAGGCTGGGAGTGACGTTTGAAGCTGCAGTCTTCTGTGATATTTCCCAAGGTCTTCT  
 TACACCTGGAACCTTGATGGACTCTGAAGGGCTCCCTGTCTCCCTCCCTGCTGTGTGG  
 ACACTCACAGACAGACCCTCATCTCCCGAGCCACACCTTGAGTATGGGAACATACAC  
 TGCCCTTGCCCAAGGTTTCAGATTGAAGGCAGTGTGGTGTACAGCAACTCTGTGTGGC  
 CTGGAGGTGCGAGCCAGGCCCTGTGAGTGTGATCTCCGAGGGCACACACCTATTCT  
 TCTCCAGGACCACCTCATCCCCATTGTCTCAGAGGGACCCAGTCTTCGACCCTGA  
 CGACCCTGGGGCGACTCTCAGCCACTCCGATGACTTCTCCAACAGGTATCACTGGGAA  
 TGCGCCACCGCTGGCTCCCCAGCACATCCCTGCTTCGACTCCTCCACTGCACACCAAC  
 TGGATGCCGCGGCTCCCACTGTTTCTTTGAGGCACAATGGCTCAGTCAGCAGCTATGA  
 TCAGTTCCTTGTGATGCTGAGGGTCTCCAGTGGTGGCCGGAACCTCTTGAGACCCGG  
 GTGTTCTGTCCCCCTACCCTGACTCGGCGTTCAGATTCGTCCACATCTCCTGGGTCA  
 GCTTTAAAGACACCTTCGTCAACTGGAATGACGAACCTCTCTTCAAGCTATGTGTGA  
 GGACTGCAGTGAAATACCGAATCTGTCTTATCCTGGGATCTCTTTTGTAGTCAATGCA  
 ACAGAAAAGAAATAGGATAGAAGTCCGTGTGAACACTTGTGAGGCACCCAGCAGAGAG  
 TGACACACTCAAGGGCTGCTTCTGACCTCAGTGTGATATGGAAGGCTCGGCCAACAC  
 CTGTGTAGGCAGTTTTATTGAGCTAAAGCCACAGTTCAGAAGGACCTGCGATGTGACA  
 CACTGCCAGAGGGCTGTGTCCGTCAACACCTCGCCTGTCTTTTGGGCCAGCTGCACA  
 AGTCATCACAGTTAAACCTGCTGCCCACTGAGCCTGGCACTGCAGATCTGATGCAAC  
 GACCACACCATTCTCACGGGAACCTTCACCCGTGACCCTTGCCCAACCTGCCACTTCA  
 GCTCCAAGGGGAACCCCCACAGAGCCATGACTGGAGTCTACTGGATTCTCTCTGCGG  
 GGGACTCTGCAGTCTGGGGGAGGCTCCAGAGGAAGGTTCACTAGACCTAGAGCCAGG  
 GCCACAGAGCAAGGGATCCCTGATGACTGGCCGCTCTGAGAGAAGTCAGCCACCCAC  
 AGCCCTGACCCTCACCTCTCTGCTAAGGACACCAGCTTTCCAGGATCAGGACCTAGCT  
 TGAGTGCCGAGGAGAGCCCTGGAGATGGGGATAACCTGGTGGACCCCTCCCTGTCTGC  
 AGGCAGAGCCGAGCCTGTCTCATGATTGACTGGCCCAAGGCCCTGCTGGGTGAGCA  
 GTTTTCCAAGGCTATTCTCTCAGGTATTACAGAACAGACAGTGACCAATCAAGCCAT  
 ACTCTCTGAGCAGTGGAGAGACGTACGTCTGCAAGTGTCTGTGGCTTCGAAGCATGG  
 CTTACTGGGTAAAGCTCAGCTGTACTTGACAGTCAACCCGCTCTCTGGGACATGGCC  
 TGTGAGGTGCAGCCCCACCATGGTCTGGAAGCACACACCGTCTTCAGTGTCTTCTGCA  
 TGTCTGGAACCCGACTTCCATTATGAATTTAGTTACCAGATAGGAAACACCTCCAA  
 ACACACTTTGTACCATGGGAGAGACACCCAGTATTATTTTGTGCTGCGAGTGTGAG  
 CACTTGGACAATTACAAAGTCATGGTTTCCACTGAAATCACAGATGGCAAAGGCTCCA  
 AGGTCCAGCCGTGCACTGTGGTGGTACTGTGTGCCCCGCTACCATGGAATGACTG  
 TCTGGGCGAGGACCTGTATAATTCCAGCCTGAAAAACCTTCTACCTCCAGCTGATG  
 GGGAGTTACACAGAAATCAGGAATACATCACTGTGATCACCAGAATCCTGAGTCGTT  
 TGTCTAAGGAGGACAAAACCTGCCTCCTGCAACCAATGGTCACGAATACAGGATGCATT  
 AATTTCTTCAGTATGCAGATTGGCTTTTGTAGATCAGTAGGCTTTATGAGTGGGTT  
 CTCATCCTCAAGTACACCCGGGCACTCCTTGCTCAAGGCCAGTCTCGGGGCCATTG  
 TGATTGACAAAGGAGTGAGGCTTGAGCTCATCGGTCTCATATCCAGAGTCTGGGAAGT  
 CTCTGAGCAAGAAAACCTGAAGGAGGAAGTCTATCGACATGAAGAAGGAATTACAGTC  
 ATCTCAGATTTATTGTTGATTGGTGGAGTTGTGGGCCTCAACCTCTATACCTGCTCCA  
 GCAGAAGACCCATCAACAGGCAATGGCTAAGGAAACCCGTGATGGTCGAGTTTGGGGA  
 GGAGGATGGCCTGGATAATAGGAGAAATAAACGACATTTGTATTACTTCGGGATAAA  
 GTGAATCTCCATCAGTTCACTGAGCTTTCCGAAAACCCCAAGGAATCTCTACAGATAG  
 AAATTGAATTTTCCAAACCTGTTACAAGGGCATTTCCCGTCATGTTGCTAGTAAGATT  
 CTCTGAGAAACCTACTCCCTCTGATTTTCTGTGAAGCAGATCTACTCTGGGATGAG  
 TCAATTGTGCGAGATTTATATACCTGCTGCTTCTCAGAAAAGATGCCAGTGTAGGCTATT  
 TATCCTTATTGGATGCTGACTATGACAGAAAACCTCCAAACAGATATTTAGCTAAGGC  
 AGTGAATATACAGTACATTTCCAGTGGATCCGATGCCTGTTTTGGGACAAGAGAGAG  
 TGGAAATCTGAACGTTTCTCTCCACAACAGGGACTTCTCCTGAAAAAGTGAACCTGCA  
 GCTACCATCGCCTCGCGGCATTGCTCTCCTAAGGAGAAAAGCTGAAGGCCAGTTTTGA  
 AGTGAGTGACATTTCCAAGCTACAGAGCCACCCAGAAAACCTTGCTTCCAGTATTTTT  
 ATTATGGGTTCTGTGATTCTTTATGGATTTTGGTTCGCTAAAAGTAGACAAGTAGATC

	<p>ATCATGAAAAAAGAAAGCTGGTTACATCTTTCTGCAAGAAGCTTCCCTGCCGGGCCA  TCAGCTATATGCGGTCGTTCATTGACACTGGCTTCCGAGCTCCGGCCAGCGCTCCTGCC  CAACTGGGCTGCTGAGGAAGATCCGCCTCTGGCAGCAGCCGTGGGCTTCCCCAG  GCTGGTTCATCAGCCACGTGATGGTGAAGGAGCTGCACACGGGACAGGGCTGGTTCTT  CCCTGCCCAGTGTGGCTGTCTGCCGGCAGGCATGATGGTCGCGTGGAGCGGGAGCTC  ACCTGTCTGCAAGGGGGACTCGGCTTCCGGAAGCTTTTCTATTGCAAGTTCACAGAGT  ACCTGGAGGATTTCCATGTCTGGCTGTGCGGTGTACAGCAGGCCCTCCTCCAGCCGTA  CCTGCACACGCCGCGCCTCACCGTGTCTTCTCCCTGCTGTGCGTCTACCGTGTCTC  ACTGCCCTGGTTGCTGCTGGAGGGCAAGAGCAGGTGAGAGCCATCGTTTTCTTATA  GCAGCTTCAGATCCGACTACACTGTGGCCCTTTTGCCTAAGAAATCAACAAAGCT  CACAGTTCTCCGAGAAAAGTTAAACCAGGGGAAGCAAGCCTGGCTGCCTGGGGACCA  GAGAAGGAACAGGAGGGCTCTGCCCGGCTCAGCAAGGTACCTGCGACTGTCTCATG  GCCCTGTTCTCCTGAGCAGCCCTTCATGTCTGGGGAACACGCTTGGCGAACACCTC  TTTCTTCTGCAGGAAGCCCCGGGTCTGCCCGAGTGGAGCCACACAGCCCACTTAGA  GGAGGAGCAGACCGAGGCCACCCATGGTGGGTGAGAAAGAGGGGTCTCAGCAGAG  GCCTGAAACAGGAAGGAAGTGAAGCCCAGAAGAATTCAGAAAGCCCTGTGTGTCTACT  CAGTAAATACCGGCAGGACCGTGGGAGAGACACTGTGGAGCAGCAAGGCTCGGGCACC  CAGCAGTGGTTTGGAGGGACTAATGCCCCAGTGGTCAAGGGCCCTTCAGCCTTGGTGG  AGCTCTGCAGTGTGGGCCATTGTGGGACCGCTTCTTGGCCTGCAGTTGGGGACAG  GATTTCTAGCCTACAGGTATGCCTCATGGCCTTGGGTTTTGCTTGGAAAAGAAGAGCT  GACAACCACTTTTTTACTGAGTCTTTATGTGAGGTACCAGGGATCTGGACTCTGAAT  TGGCAGAACGTTCTTGACTCGCCTCCCTTCTCTTCAAGCTGCAGTATTCCTGACTG  TGCAGGCGAGGTTGAAAAAGTCTTGGCTGCCCGACAACAAGCTCGCCACCTGCGCTGG  GCGCATCCACCATCCAAGGCCAGCTGAGGGGCACCCAGACAGAGGATGAGGAGAGAGA  GTCGCACACGGGCTGCCCTGAGAGACATTTCCATGGACATCCTCATGCTGCTTCTGCT  TTTGTGTGTAATATATGGGAGATTTTCCAAGATGAATACTCCTCAATCAAGCTATC  CGGAAAGAAATTTACAAGAAATGCCAGAACTGCTTGGGTGGCCTGAGAAACATCGCTG  ACTGGTGGGACTGGAGTCTGACCACACTTCTGGATGGCCTGTACCCGGGAGGCACCCC  GTCAGCCGCTGTGCCGGGGGCTCAGCCTGGAGCTCTTGGAGGAAAATGCTACCTAATA  GGCAGTTCGTAATTAGGCAGCTAAAAGTTTTTCTAGGCATTTATGCAAGCCTCCCA  GGCCATTTTCAGCACTCATCGAAGACTCTATTCTACATGTAGTCCCGAAGTTGGAGG  CCCTGAGAACCCCTACCTGATAGACCCAGAGAACCAAAACGTACCCTGAATGGTCTCT  GGGGGCTGTGGGACAAGGGAGGACTGTGTGCTCAGCCTGGGCAGAACAAAGGACTGAAG  CCACACAGCCCTGTCCCGACTCAGGGCCAGCATGTGGATTGACCGCAGCACCAGGC  TGTGTCTGTGCACTTCACTCTCTATAACCCCTCAACCCCACTCTTACCACGCTGTCC  CTGAGAGTGGAGATCCTCCCTACGGGGAGTCTCGTCCCTCATCCCTGGTGGAGTCAT  TCAGCATCTTCCGAGCGACTCAGCCCTGCAGTACCACCTCATGCTTCCCCAGGTGAG  CTGACCTGCCTCTTGGGCCTCCTGGAGGTGCACAGGAAGATGGGGCTTACCTGGGCT  GGGCTTCTCCACCAGACAGGACTAGTTCCTTACCCAT</p>
	<div>ORF Start: ATG at 1</div> <div>ORF Stop: TGA at 6904</div>
	<div>SEQ ID NO: 138</div> <div>2301 aa</div> <div>MW at 254558.5kd</div>
NOV29a, CG59297-01 Protein Sequence	<p>MGWRGLIAALPLLSLVQPALGTSSKDEDVGRSWSADCHTCDQLAQDMAEAAQNI SDD  QERCLQAACCLSFGGELSVSTDKSWGLHLCSCSPGGGLWVEVYANHVLMSDGKCGC  PWCALNGKAEDRESQSPSSASRQKNWKTTSEAALSVVNEKTQAVVNEKTQAPLDCD  NSADSLRVFADSSIGENWTLQMVCDPDTWMRGPPSSHGLPPGIPRTPSFSTASQSGSEIL  YPPTQHPPVAILARNSDNFMNPVLNCSLEVEARAPPNLGFRVHMASGEALCLMDFGD  SSGVEMRLHNMSEAMAVTAYHQYSKEGVYMLKAVIYNEFHGTEVELGPYYVEIGHEAV  SAFMNSSSVHEDEVLFVADSQVNQKTVSVYTNGTVFATDITFTAVTKETI PLEFEW  YFGEDPPVRTTSRSIKKRLSIPQWYRVMVKASNRMSVSVSEPHVIRVQKKIVANRLTS  PSSALVNASVAFECWINFGTDVAYLWDFDGTVSLGSSSSSHVYSREGEFTVEVLAFN  NVSASTLRQQLFIVCEPCQPPLVKNMGPVKVQIWRSPVRLGVTFEAAVFCDISQGLS  YTWNLMDSIEGLPVSLPAAVDTHRQTLILPSHTLEYGNYTALAKVQIEGSVVYSNYCVG  LEVRAQAPVSVISEGTHLFFSRTTSSPIVLRGTQSFDPDDPGATLSHSDDFSRYHWE  CATAGSPAHPFCFDSSTAHLQDAAAPTVSFEAQWLSYDQFLVMLRVSSGGRNSSETR  VFLSPYPDSAFRFVHISWVSFKDTFVNWNDELSQLAMCEDCSEIPNLSYSWDLFLVNA  TEKNRIEVRVNTCEAPAEVTHSRAASDLVSIWKAAPNTCVGSFIELKPQFRRTCDVT  HCEGECVRHHLACLGLQHLKSSQLNLPTPEPGTADPDATTTTSPSREPSPTLGLQPATS  APRGTPTEPMTGVYWIIPAGDSAVLGEAPEEGSLDLEPGPQSKGSLMTGRSERSQPTH  SPDPHLSAKDTSFPGSGPSLSAEESPGDGDNLVDPPLSAGRAEPVLMIDWPKALLGRA</p>

	VFQGYSSSGITEQTVTIKPYSLSSGETYVLQVSVASKHGLLGKAQLYLTVPAPRDMA CQVQPHHGLEAHTVFSVFCMSGKPDFHYEFSYQIGNTSKHTLYHGRDTQYYFVLPAGE HLDNYKVMVSTEITDGKGSKVQPCTVVVTVLPRYHGNDCLGEDLYNSSLKNLSTLQLM GSYTEIRNYITVITRILSRLSKEDKTASCNQWSRIQDALISSVCRLAFVDQLGFMSAV LILKYTRALLAQGFSGPFFVIDKGVRELIIGLISRVEVSEQENSKEEVYRHEEGITV ISDLLLIGGVVGLNLYTCSSRRPINRQWLKPKVMVEFGEDGLDNRNKTTFVLLRDK VNLHQFTELSNPQESLQIEIEFSKPVTRAFFVMLLVRFSEKPTPSDFLVKQIYFWDE SIVQIYIPAASQKDASVGYSLLDADYDRKPPNRYLAKAVNYTVHFQWIRCLFWDKRE WKSERFSPQPGTSPEKVNCSYHRLAALFALLRRKLKASFEVSDISKLSHPENLLPSIF IMGSVILYGFLVAKSRQVDHHEKKKAGYIFLQEASLPGHQLYAVVIDTGFRAPASAPA QLGLLRKIRLWHSRGPSPGWFISHVMVKELHTGQGWFFPAQCWLSAGRHDGRVEREL TCLQGGLGFRKLFYCKFTEYLEDHFWLSVYSRPSRRYLHTPRLTVSFSLLCVYACL TALVAAGGQEQVRAIAFFPYSSFQIRLHCGPFLPKKSTKLTVLREKFKPGEASLAAWGP EKEQEGSARLSKVPATCPHGVLSSPFIAGEHAWRTTSFLLQEAAPGSARVEPHSPLR GGAQTEAPHGGSEERRGLSRGLKQEGSEAQKNSESPVCLLSKYRQDRGRDTVEQQSGT QQWFGGTNAPVVKGPSALVELCSVGHLDWDRFFGLQFGDRISLQVCLMALGFAWKRA DNHFFTESLCEATRDLDSELAERSWTRLFFSSCSIPDCAGEVEKVLAAQQARHLRW AHPPSKAQLRGTRQMRRESRTAALRDISMDILMLLLLLLCVIYGRFSQDEYSLNQAI RKEFTRNARNCLGGLRNADWWDWSLTLLDGLYPGGTPSARVPGAQPGALGGKCYLI GSSVIRQLKVFPRLCKPPRPFSAIEDSIPTCSPEVGGPENPYLIDPENQNVTLNGP GGCGTREDCVLSLGRTRTEAHTALSRLRASMWIDRSTRAVSVHFTLYNPFTQLFTSVS LRVEILPTGSLVPSSSLVESFSIFRSDSALQYHMLLPQVS
--	---

Further analysis of the NOV29a protein yielded the following properties shown in Table 29B.

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 22 and 23

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 29C.

Table 29C. Geneseq Results for NOV29a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU14647	Novel bone marrow polypeptide #46 - Homo sapiens, 488 aa. [WO200157187-A2, 09-AUG-2001]	196..683 1..488	487/488 (99%) 487/488 (99%)	0.0
AAU29269	Human PRO polypeptide sequence #246 - Homo sapiens, 300 aa. [WO200168848-A2, 20-SEP-2001]	2046..2300 1..255	254/255 (99%) 255/255 (99%)	e-147
AAE03429				e-139



	protein HETDB76, SEQ ID NO: 112 - Homo sapiens, 561 aa. [WO200132675-A1, 10-MAY-2001]	1..240	240/240 (99%)	
AAU14741	Novel bone marrow polypeptide #140 - Homo sapiens, 142 aa. [WO200157187-A2, 09-AUG-2001]	478..618 2..142	140/141 (99%) 140/141 (99%)	5e-79
AAB41274	Human ORFX ORF1038 polypeptide sequence SEQ ID NO:2076 - Homo sapiens, 160 aa. [WO200058473-A2, 05-OCT-2000]	1621..1738 43..160	116/118 (98%) 116/118 (98%)	5e-66

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29D.

**Table 29D. Public BLASTP Results for NOV29a**

Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96Q08	KIAA1879 PROTEIN - Homo sapiens (Human), 995 aa (fragment).	1449..1734 212..538	77/332 (23%) 136/332 (40%)	5e-17
CAB59175	SEQUENCE 3 FROM PATENT WO9518225 - Homo sapiens (Human), 1614 aa (fragment).	1621..1737 403..523	45/121 (37%) 67/121 (55%)	2e-16
CAB59174	SEQUENCE 1 FROM PATENT WO9518225 - Homo sapiens (Human), 4339 aa (fragment).	1621..1737 3128..3248	45/121 (37%) 67/121 (55%)	2e-16
O42181	PKD1 PROTEIN - Fugu rubripes (Japanese pufferfish) (Takifugu rubripes), 4578 aa.	308..795 2000..2473	120/517 (23%) 191/517 (36%)	2e-16
Q15141	POLYCYSTIC KIDNEY DISEASE 1 PROTEIN - Homo sapiens (Human), 4292 aa.	1621..1737 3161..3281	45/121 (37%) 67/121 (55%)	2e-16

Pfam analysis predicts that the NOV29a protein contains the domains shown in the Table 29E.

**Table 29E. Domain Analysis of NOV29a**

Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PKD: domain 1 of 2	371..454		0.14

		50/94 (53%)	
PKD: domain 2 of 2	456..536	24/93 (26%) 61/93 (66%)	3.7e-09
REJ: domain 1 of 1	592..710	39/144 (27%) 74/144 (51%)	0.0013
hormone3: domain 1 of 1	1221..1233	6/13 (46%) 13/13 (100%)	7.8
GPS: domain 1 of 1	1497..1544	13/54 (24%) 31/54 (57%)	0.18
PLAT: domain 1 of 1	1623..1684	15/69 (22%) 46/69 (67%)	2e-05

**EXAMPLE 30.**

The NOV30 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis		
	SEQ ID NO: 139	3095 bp
NOV30a, CG59264-01 DNA Sequence	CTGGCCAGACCCTGCCTCCAGCCACCGAGGCACATTACCTGGGCCCAACAGATGTCCTT GGCAGCGGGGCCACAGTATCCTCCTCATGCGTCCACTCTCTTCAAGCCAGTGCAAAG AAAGCAGCGAATCCAGCACCTGAAAAATACCCAGGAGGTCTCGAACCCTCTGGT AACTCTCAACCCCTCATCTGCGAGCACGTGGGGCTTGGGGCCACGTTTGTATGTTGGG GAGGGCCCTGTGTTTGGGGGAATGCAGCTCCCTTCTGCAGAGATGGGAGGTGGTCAT GAAGCTGGACCGCAAGGCCTTGGCTCGCGATCCGCCCGCTTCAGCCTCCCAAAGTACT GGGATTACAAGAGTGAGCCACTGCGCCAGGCCTCAAACATCAATGTTGATACCTGTTT TCAAAGTACTGGAAGAAGGTAGGGGTAAAGGACAAGGAACCGGAGGAGTGGAGGGCGT CACTGGGTTCGGCGTCTGGCAAGCGGTTCACTGTCTGCTCCCTAGCAGCCGGCCTT CGGGTCGGGCGTCTCGCCGCTACTGCGGCTTCAGTTCTCCCGGTGTGGCCACGAGT CGGGTGAGTCTCGGTTCAAGACACTCAGCCGGGGATGCCAGAGCCTGTGGGCAGCCG TATCGAACACGCGAGGTATCGTGGCAATCCAGAACGCTTTTCTGAATGGGATAGTTTG AAAGAAGGGCAGGCGTCTTTGTGGCACGGTAGGAACCTGGGATGCACCTTCGCCGCCCA GAGAGAAATACATAAAATCTATTGAGCGAGCGCTTGTTAATTATATGTGCCTTCTGCT TATTATATGCCTACAGGTACAGGAACCTCAAGTATTTACAGAATGATGCCAAGGAG CGCAGTTTCTACACCATCCCTCCCGATCAGACCACATCTATTCCAATCTCAGTCTCAA ACTCTTCCCCTTTTCAACTCACATTAATGAATTAAGTACCTAGCACTGAGAAT AAAAAGCGAAAGCAGCATCCGTCTTCTACTCTCACACAGTTACAGTCCACAGCAA AGCGCAAGTCATCAAACACACCTGGTTGGCAGTGCCAGATTCTGCAAGTGAGGTCC AGGAAAGCTCTTGCCCTCTTGCCAGCAGCCGAGTATCTCAACGGATGCCGTGCACC ATATTCCCTGGATGCTGAAGACATGGCAGACTATGGGTGGCAGTACCAGAGCCAGGAC CAACGTCAAGGGTATCCCATCTGGGGCAAACCTCACTGTGTACCGGGGAGGAGGCTACG TGGTCCCCTTGTCCAGGACTAGGCAACAAGAGCGAAACTCTGTCCCTGGCAAAAAA GAACACCTGGCTGGACGCCCTGACCAGAGCTGTGTTGTGGAGTCCACTGTCTACAAC GCCAACGTCAACCTGTTCTGCATTGTACGCTGACGCTAGAGACCAGCGCTCTGGGTG GGTATTTTGAATTTCTTTTCAAGAAATCATAAATTTCTATCTAACTTTGGGGTCATT CGTGGTAGCGGCAGAGCTCATCTACTTCTCTTCTCTCTACTACATTGTGGTGCAA GTGCTTGAATCCAGGAGGCACAGGTTGCACTATTTCTGCAGCAAGTGGAACCTTCTGG AGCTGGCCATCATCTGGCCAGCTGGAGCGCCCTGGCGGTGTTGTGAAGAGGGCTGT CCTGGCCGAAAGGGACCTCCAGATCATTGAGACTGAGGGCGCTCTACCGAACTTCCAA GCTGTTCAAGGATCAACTATACAAATGAACAAATTATCCGCCTTCTGTGTAATCCTGT CCACAGTGAAGCTTTGGCATCTGCTCAGGTTGAATCCCAAAATGAACATGATCACGGC AGCCCTACGCCGTGCCTGGGGCGACATTCAGGCTTTATGATTGTATCCTTACCATG CTCCTGGCTTACTCCATCGCGTAAGTATCTGCTTGGGTGGAACCTCCGTTCTACA	

	AAACCCCTTTGATGCGGCGGAGACGATGGTCAGCCTTCAGCTGGGAATCTTCAACTATCGAGGAGGTCCTGGACTATAGCCCAAGTGCTTGGCTCCTTCTCTATTGGATCCCCACTGCACCTGGCCACATTTCTGTTTTTTTTTTTTTTTTTTTTTTTGGATGGAGTATGGCTATTGCATCACACAGACAACCTTCATTAAAGGAAGACACCAAGACAGGAGCTGCTCAGGGGCCACTGGGCACTGCGGTTAGAAAGGGAGCGAGACGCTCTCTCAAAAAAAAAAGAAAGAAAAAATATTCTGGAAGAGGCAGGAGAATCACTTGAATCCGGGAGAGGGAGGTTGAGTTTCGAAAGGAAGCGAGAGGGAGCGAAAGGCAGAGGCACTATGTGCCGGGGCGTCACTCTGCTTGTCAAGAAAAAATGATGGATGGGAAGAAAAAGATCACACATCAACATCAACATATGAAAAGCATGAGAGACAATTTAAAAAAGAGGACCTCCCCTCACTGCATTTTCCCTCACCACAACCCCTCACTCCCAAGACTTTCCCAAAAGATCAGGAACTAAGCCTGAGAGAAGCCAATGTAGGACAACGTGGGTACGGGAGCCGGGAGCGGGCACTTGAAGCGGTAGAGCCAAGGAGCCGGTTTGAAGTTTGTCTGAAGGTCATGAGAGGCGACCGAAGGATTTTAAGTGTAGGGGAGAAGAAGTTGAGGAGCCGGGATATCATTGGAAGGATCTTGGAGTCAGTGGCAGGCAAACAAAATTTAGGTAAGGACTCACAGGAAGAGGTTGAGGCATGAGGAGGAATTATCAAGAGTGAAGTGCACAGATGGGAAACAGCCCAAGGAACACCCCTTGTGACTGTAGCATGCGGTACAAAGGGCCCTGAGTGTCCAGCCTAAGCAACAGAGCAAGACTCAGTCTCAAAAAAACAACAAAAAATCCCTGGGCGTGGTGGCTCATGCCTGTAATCTCAACACTTTGGGAGAAAAATATATATATTTTTCCCTTAAATTATCATGTTGCAGGCCGGGCACAGTGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCCAAGGCAGCGGATCACCTGACGTAAGGAG		
	ORF Start: ATG at 52	ORF Stop: TAA at 2959	
	SEQ ID NO: 140	969 aa	MW at 108791.3kD
NOV30a, CG59264-01 Protein Sequence	MSWQRGHSILLMRPLSSSPVQRKQRI PAPEKYPGGPRTSHSGNSQPLICEHVGLGATFVCWGGPCVLGECSSLLQRWEVVMKLDRLKALARDPPASASQSTGITRVSHCARPQTSMLIPVFKVLEEGRGKDKPEPGVEGVTGFRRLASGSVCSLAAGLRVGLRLRLPLQFSRCGHESGESRFEDTQPGMPEPVGSRIEHAGYRGNPERFSEWDSLKEGQASLWHRNWDALSPPREKYIKSIERALVNYMCLLLIICLQVTGTSSISQNDAKERSFTYIIPDQTTISIPISVSNSSPFSTHINEFTKYLALRIKSESSIRLLYSHYTAISPQQSASHQTHLVGSAQIRQVRVQESSCLPAQQPQYLNCRAPYSLDAEDMADYGWQYQSQDQROGYPYIWGKLTVYRGGGYVVPVLSRTRQQRNSVPVGGKKNTWLDALTRAVFVESTVYNANVNLFICIVTLTLETSALGGYFEFLFKKFINFYLTLSFVVAELIYFLFLYYIIVVQVLESRRRLHYFCSKWNLLELAIILASWSALAVFVKRAVLAERDLQIIETEGALPNFQAVQGSTIQMNKLSAFLVLLSTVKLWHLRLNPKMNMITAAALRRAWGDISGFMIVILTMLLAYSIAVSICFGWKLFSYKTLFDAETMVSLQLGIFNYEEVLDYSPVLGSFLIGSPLHLATFLFFFFFFFLRWSMAIASHRQPSLKEDTKTGAAQGPLGTAVRKGARRSLKKRKKKNILEEAGESLESGRGRLQFERKREGAKRGRTMCRGVSLLVKKKMDGKKKITHQHQHMKSMRDNLKKEDLPSEHFPSPQPLTPKTFPKDQETKPKERSQCEDNVGHGSRERALEAGRATGAGLSFVVRVMRGDRRILSVGEKKLRSRDIIGRILESVAGKQNLGKDSQEEVRQMEELSKSEVHRWETAHFNNTVVTVACGKTGKPECPKQKQSKTQSQKKQTKKIPGRGGSC		

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3869 probability located in mitochondrial inner membrane; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 30C.

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68450	Amino acid sequence of a human PKD2 polypeptide - Homo sapiens, 968 aa. [US6228591-B1, 08-MAY-2001]	15..967 2..961	322/1001 (32%) 493/1001 (49%)	e-125
AAY78946	Polycystic kidney disease PKD2 amino acid sequence - Homo sapiens, 968 aa. [US6031088-A, 29-FEB-2000]	15..967 2..961	322/1001 (32%) 493/1001 (49%)	e-125
AAM51861	Murine polycystic kidney disease protein 2 - Mus musculus, 966 aa. [WO200177331-A1, 18-OCT-2001]	63..967 39..959	302/960 (31%) 467/960 (48%)	e-116
AAB68448	Amino acid sequence of an internal fragment of human PKD2 - Homo sapiens, 866 aa. [US6228591-B1, 08-MAY-2001]	423..967 295..859	188/580 (32%) 303/580 (51%)	2e-76
AAY70245	Human Polycystin-L protein - Homo sapiens, 805 aa. [WO200012046-A2, 09-MAR-2000]	217..807 75..668	164/616 (26%) 289/616 (46%)	2e-60

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a				
Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13563	Polycystin 2 (Autosomal dominant polycystic kidney disease type II protein) (Polycystin) (R48321) - Homo sapiens (Human), 968 aa.	15..967 2..961	320/1001 (31%) 490/1001 (47%)	e-123
O35245	Polycystin 2 - Mus musculus (Mouse), 966 aa.	63..925 39..916	295/917 (32%) 454/917 (49%)	e-114
G02640	polycystic kidney disease protein 2 - human, 608 aa (fragment).	383..967 6..601	202/611 (33%) 321/611 (52%)	6e-92
Q9UP35	POLYCYSTIN-L - Homo sapiens (Human), 805 aa.	217..807 75..668	165/616 (26%) 290/616 (46%)	3e-60
Q9P0L9				3e-60

	2-LIKE PROTEIN - Homo sapiens (Human), 805 aa.	75..668	290/616 (46%)	
--	---	---------	---------------	--

PFam analysis predicts that the NOV30a protein contains the domains shown in the Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	NOV30a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ion_trans: domain 1 of 1	489..688	41/233 (18%) 139/233 (60%)	2.4e-06

#### EXAMPLE 31.

The NOV31 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis		
	SEQ ID NO: 141	2316 bp
NOV31a, CG59623-01 DNA Sequence	CCTGAGCCTCATTGGGGGGTCTCCCCCACGGGCCGGGCATGCTGCCCCCGGAAG GAACCCCTCTCCTCGCTCCCCCAGCGTCCACGCGGAGCATGAACATTGAGGATGGCG CGTGCCCGGGCTCCCCGTGCCCCCGCTGCCGCCGGTAGGATGTCTGGCCCCACG GGGCATGTCTCTCTCTGCTCTTCTCCCCACCCTGGGGGCCGGTGGAGGTGGAGT GGCCGTGACGTCTGCCGCCGAGGGGGCTCCCCGCCGGCCACCTCTGCCCCGTGGCC TGCTCTGCAGCAACCAGGCCAGCCGGTGATCTGCACAGGAGAGACTGGCCGAGG TCCAGCCAGCATCCCGGTCAACACGCGGTACCTGAACCTGCAAGAGAACGGCATCCA GGTGATCCGGACGGACACGTTCAAGCACCTGCGGCACCTGGAGATTCTGCAGCTGAGC AAGAACCTGGTGCACAAGATCGAGGTGGGCGCCTTCAACGGGCTGCCAGCCTCAACA CGCTGGAGCTTTTGGACAACCGGCTGACCACGGTGCCACGCAGGCCTTCGAGTACCT GTCCAAGCTGCGGGAGCTCTGGCTGCGGAACAACCCCATCGAGAGCATCCCCTCTAC GCCTTCAACCGCTGCCCTCGCTGCGGCGCCTGGACCTGGGCGAGCTCAAGCGGCTGG AATACATCTCGGAGGCGGCCTTCGAGGGGCTGGTCAACCTGCGCTACCTCAACCTGGG CATGTCAACCTCAAGGACATCCCCAACCTGACGGCCCTGGTGGCCTGGAGGAGCTG GAGCTGTGGGCAACCGGCTGGACCTGATCCGCCCGGGCTCTTCCAGGGTCTACCA GCCTGCGCAAGCTGTGGCTCATGCACGCCAGGTAGCCACCATCGAGCGCAACGCCCT CGACGACCTCAAGTCGCTGGAGGAGTCAACCTGTCCACAACAACCTGATGTCGCTG CCCCACGACCTCTTACGCCCCCTGCACCGCCTCGAGCGCGTGCACCTCAACCACAACC CCTGGCATTGCAACTGCGACGTGCTCTGGCTGAGCTGGTGGCTCAAGGAGACGGTGCC CAGCAACACGACGTGCTGCGCCCGCTGTCATGCGCCCGCCGGCCTCAAGGGGCGTAC ATTGGGGAGCTGGACAGTCGCAATTCACCTGCTATGCGCCCGTATCGTGGAGCCGC CCACGGACCTCAACGTACCGAGGGCATGGCTGCCGAGCTCAAATGCCGCACGGGCAC CTCCATGACCTCCGTCAACTGGCTGACGCCCAACGGCACCTCATGACCCACGGCTCC TACCGCGTGCGCATCTCCGTCTGTCATGACGGCACGCTTAACTTCACCAACGTACCG TGCAGGACACGGGCCAGTACAGTGCATGGTGACGAACTAGCCGGCAACACCACCGC CTCGGCCACGCTCAACGTCTCGGCCGTGGACCCCGTGGCGGCCGGGGGACCGGCAGC GGCGGGGGCGGCCCTGGGGGAGTGGTGGTGTGGAGGGGGCAGTGGCGGCTACACCT ACTTCAACACGGTGACCGTGGAGACCTGGAGACGACGCCGGAGAGGAGGCGCTGCA GCCGCGGGGACGAGAAAGGAACCGCCAGGGCCACGACAGACGGTGTCTGGGGTGGG GGCCGGCCTGGGGACGCGGCCGCCCTGCCTCGTCTTCTACCACGGCACCCGCCCGC GCTCCTCGCGGCCACGGAGAAGGCGTTACGGTGCCCATACGGATGTGACGGAGAA CGCCCTCAAGGACCTGGACGACGTATGAAGACCACCAAAATCATCATCGGCTGCTTC GTGGCATCACGTTTCATGGCCGCGGTGATGCTCGTGGCCTTCTACAAGCTGCGCAAGC AGCACCAGCTCCACAAGCACCACGGGCCACGCGACCGTGGAGATCATCAACGTGGA	

	GGACGAGCTGCCCGCCGCTCGGCCGTGTCGTTGGCCGCCGCCGCCGCGCCGCTGGCCAGT GGGGGTGGTGTGGGCGGGGACGCCACCTGGCCCTGCCCGCCCTGGAGCGAGACCACC TCAACCACCACCTACGTGGCTGCCGCCTTCAAGGCGCACTACAGCAGCAACCCAG CGGCGGGGCTGCGGGGGCAAAGGCCGCTGGCCTCAACTCCATCCACGAACCTCTG CTCTTCAAGAGCGGCTCCAAGGAGAAGTGCAAGAGACGAGATCTGAGGCGGGGGG CCGGCGGGCGAGGGGCGTGGAGCCCCCACCAGGTCCCAGCCGGGCGCAGC		
	ORF Start: ATG at 111	ORF Stop: TGA at 2250	
	SEQ ID NO: 142	713 aa	MW at 76433.0kD
NOV31a, CG59623-01 Protein Sequence	MARARGSPCPPLPPGRMSWPHGALLFLWLFSPPLGAGGGVAVTSAAGGSPPATSCP VACSCSNQASRVICTRRDLAEVPASIPVNTRYLNLQENGIQVIRTDTFKHLRHLEILQ LSKNLVRKIEVGAFNGLPSLNTLELFDNRLTTVPTQAFEYLSKLRELWLRNNPIESIP SYAFNRVPSLRRLDLGELKRLEYISEAAFEGLVNLRYLNLGMCNLKDI PNLTALVRLE ELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNAFDDLKSLEELNLSHNNLM SLPHDLFTPLHRLERVHLNHNPNWHCNCVWLSWVKETVPSNTTCCARCHAPAGLKG RYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAELKCRGTGTSMTSVNWLTPNGTLMTH GSYRVRISVLHDGTLNFTNVTVDGTGQYTCMVTNSAGNTTASATLNVSAVDPAAGGT GSGGGGPGSGGGVGGGGGGYFTTVTVETLETQPGEEALQPRGTEKEPPGPTTDGVW GGGRPGDAAGPASSSTTAPAPRSSRPTKAFVTPITDVTENALKDLDDVMKTKIIG CFVAITFMAAVMLVAFYKLRQQLHKKHGPTRTVEIINVEDELPAASAVSVAAAAAV ASGGGVGDSHLALPALERDLNHHHYVAAAFKAHYSSNPSGGGCGGKGPGLNSIHE PLLFKSGSKENVQETQI		

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Likely cleavage site between residues 38 and 39

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 31C.

Table 31C. Geneseq Results for NOV31a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE13006	Human leucine-rich repeat (LRR) family member protein - Homo sapiens, 713 aa. [WO200175105-A2, 11-OCT-2001]	1..713 1..713	712/713 (99%) 712/713 (99%)	0.0
AAU12355	Human PRO331 polypeptide sequence - Homo sapiens, 640 aa. [WO200140466-A2, 07-JUN-2001]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0

AAU00826	Human immune response protein PRO331 (UNQ292) - Homo sapiens, 640 aa. [WO200119991-A1, 22-MAR-2001]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
AAB53089	Human angiogenesis-associated protein PRO331, SEQ ID NO:137 - Homo sapiens, 640 aa. [WO200053753-A2, 14-SEP-2000]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
AAB65292	Human PRO331 protein sequence SEQ ID NO:501 - Homo sapiens, 640 aa. [WO200073454-A1, 07-DEC-2000]	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD10322	SEQUENCE 1 FROM PATENT WO0175105 - Homo sapiens (Human), 713 aa.	1..713 1..713	712/713 (99%) 712/713 (99%)	0.0
Q9NT99	HYPOTHETICAL 45.1 KDA PROTEIN - Homo sapiens (Human), 422 aa (fragment).	216..637 1..422	422/422 (100%) 422/422 (100%)	0.0
T46266	hypothetical protein DKFZp761A179.1 - human, 421 aa (fragment).	216..636 1..421	421/421 (100%) 421/421 (100%)	0.0
Q9HCJ2	KIAA1580 PROTEIN - Homo sapiens (Human), 640 aa (fragment).	54..713 44..640	406/660 (61%) 485/660 (72%)	0.0
Q9HBW1	BRAIN TUMOR ASSOCIATED PROTEIN NAG14 - Homo sapiens (Human), 653 aa.	42..713 34..653	381/672 (56%) 475/672 (69%)	0.0

PFam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value

LRRNT: domain 1 of 1	56..85	13/31 (42%) 21/31 (68%)	4.8e-06
LRR: domain 1 of 9	87..110	6/25 (24%) 18/25 (72%)	1.1
LRR: domain 2 of 9	111..134	9/25 (36%) 17/25 (68%)	0.38
LRR: domain 3 of 9	135..158	8/25 (32%) 19/25 (76%)	0.074
LRR: domain 4 of 9	159..182	10/25 (40%) 18/25 (72%)	0.013
LRR: domain 5 of 9	183..207	7/26 (27%) 19/26 (73%)	42
LRR: domain 6 of 9	208..229	8/25 (32%) 17/25 (68%)	1.5
LRR: domain 7 of 9	230..253	12/25 (48%) 20/25 (80%)	0.0068
LRR: domain 8 of 9	254..277	5/25 (20%) 16/25 (64%)	70
LRR: domain 9 of 9	278..301	14/25 (56%) 20/25 (80%)	0.00088
LRRCT: domain 1 of 1	311..362	19/54 (35%) 36/54 (67%)	6.2e-05
ig: domain 1 of 1	378..438	15/65 (23%) 41/65 (63%)	2.2e-07

**EXAMPLE 32.**

The NOV32 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis		
	SEQ ID NO: 143	1206 bp
NOV32a, CG59247-01 DNA Sequence	AAACTGACCACAAAAGGGCTAACGGAATTTTAGGGGATGATAAATATGTTCAACATC TTGACCATGGTGGCTGGCTGTGCCCTGGCCCTGGTGATGGTTGTCCAGCTGGGGCAGC AGATATTAATGTGCCAGGCAGTGTGGCAGGTGAAGCCCCGAGTGGACCTGTAGATC GGATGGAGACCACGTGGAGTACCACTACAGCAAGGCCATGCCACTCATCTTCATGAGC AGTATGCTCTGCAGCGGCACCATGCTGATACTCACTGGGCTGGATGCCACCTTGAGG TGCACCGCAGCAAGGAGACGCACATCATCCCTTGTGTGCTGGCCATGCACAGGCCTG GTCCAAGTCTGGCCACAAGAACTACAGCTGGACAAGGCGGGCGTGACTGACGAGGTG CTGGACATTGCCATGCAGGCCTTCATCCTGGAGGTGATCTCTAAGCAAAGGGAGCCAG CCCATGTGCTCTCCAACAAGGACCACTTCAGACTCAAGTCCCTTGAATCACTGGTCTA CCTGTACACCTGTTCTCCAGCTCCAAGTTCCTGCTTATGGCTCAGGACAGCCATGTC TCCATGCACTCCTTGATCACATGCAAGGTCACTATTGCAGGCTTCGACCTCAGCAGCT ATGGCACTGCCTACCAAGTGAACAAGGCCATAGAGGTGATGTACACCCAGTGCAT GGAGGTGGGCAAGGACAAGTGCTGCTCGTGTACTACAAGGAAGTGGTGTGCCTAGG AGCTTCCTCAGACTCATCCAGACCATCTCGGCATCACCTGGAGCAACACTGTCTCTCC	



	ACCATCAAGACCTCACTGGCAAGTGGAAATGGCATCTCCCTGTCTAAGATCCAGTGGTC CATGGATGAGGTCATCAAGCCTGTGAACCTGGAAGTGCTCTCCAAGTGACTCACCAC ATCCCTGGGGACATGGTGCCAGACATGGCCCAGATTGTCCCATGCTGGCTCAGCTTAG CCATGACCCCTATGCAAATACCCTCCCCAACCCCACTTCCACTATAGCAACCCTGAC CCCCATCATCATCAGTAACGCACACCAAGTAAGGGACTATAAAACACCAGTCAATCTG AAAGGATATTTTCAGGTGAACCAGAATAGCACCTCCTCCCACTTAGGAAGCTCATGAT TTCCAGATCTCTGCAAATGGCTTTGTTGCCCAAAAGAGAAGAACT		
	ORF Start: ATG at 38	ORF Stop: TGA at 1157	
	SEQ ID NO: 144	373 aa	MW at 41568.3kD
NOV32a, CG59247-01 Protein Sequence	MINMFNILTMVAGCALALVMVQLGQQILMCQAVLAGEAPSGPCRSDGDHVEYHYSKA MPLIFMSSMLCSGTMILITGLDAHLEVHRSKETHIIPCVLAMHQAWSKSGHKKLQLDK AGVTDEVLDIAMQAFILEVISKQREPAHVLSNKDHFRLKSLESLVYLSHLFSSSKFLL MAQDSHVSMSHLITCKVTIAGFDLSSYGNCCLKWNAIEVMYTQCMVEVGDKCLLVYY KELVLP RSFLRLIPDHLGITWSNTVLHHQDLTGKWNGISLSKIQWSMDEVIKPVNLEV LSKWTHHIPGDMVPDMAQIVPCWLSLAMTPMQIPSPPTSTIATLTPIIISNAHQVRD YKTPVNLKGYFQVNQNSTSSHLGSS		

Further analysis of the NOV32a protein yielded the following properties shown in Table 32B.

Table 32B. Protein Sequence Properties NOV32a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1279 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 28 and 29

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several

5 homologous proteins shown in Table 32C.

Table 32C. Geneseq Results for NOV32a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93565	Human polypeptide, SEQ ID NO: 3341 - Homo sapiens, 377 aa. [EP1130094-A2, 05-SEP-2001]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAM93219	Human polypeptide, SEQ ID NO: 2626 - Homo sapiens, 377 aa. [EP1130094-A2, 05-SEP-2001]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAY69421	Amino acid sequence of human TPST-2 polypeptide - Homo sapiens, 377 aa. [WO9965712-A2, 23-DEC-1999]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123

AAAY84306	A human tyrosylprotein sulfotransferase 2 (TPST-2) polypeptide - Homo sapiens, 377 aa. [WO200014250-A1, 16-MAR-2000]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123
AAAY06625	Human tyrosylprotein sulfotransferase TPST-2 - Homo sapiens, 377 aa. [WO9938980-A2, 05-AUG-1999]	10..373 10..377	240/376 (63%) 273/376 (71%)	e-123

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32D.

Table 32D. Public BLASTP Results for NOV32a				
Protein Accession Number	Protein/Organism/Length	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60704	Protein-tyrosine sulfotransferase 2 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-2) (TPST-2) - Homo sapiens (Human), 377 aa.	10..373 10..377	240/376 (63%) 273/376 (71%)	e-122
O88856	Protein-tyrosine sulfotransferase 2 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-2) (TPST-2) - Mus musculus (Mouse), 376 aa.	10..373 10..376	237/375 (63%) 270/375 (71%)	e-121
O70281	Protein-tyrosine sulfotransferase 1 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-1) (TPST-1) - Mus musculus (Mouse), 370 aa.	10..324 10..332	159/326 (48%) 213/326 (64%)	2e-78
O60507	Protein-tyrosine sulfotransferase 1 (EC 2.8.2.20) (Tyrosylprotein sulfotransferase-1) (TPST-1) - Homo sapiens (Human), 370 aa.	10..363 10..368	168/369 (45%) 226/369 (60%)	2e-78
Q9VYB7	Probable protein-tyrosine sulfotransferase (EC 2.8.2.20) (Tyrosylprotein sulfotransferase) (TPST) - Drosophila melanogaster (Fruit fly), 385 aa.	46..324 57..333	131/280 (46%) 182/280 (64%)	5e-68

PFam analysis predicts that the NOV32a protein contains the domains shown in the Table 32E.

Table 32E. Domain Analysis of NOV32a			
Pfam Domain	NOV32a Match Region	Identities/	Expect Value

		<b>for the Matched Region</b>	
Sulfotransfer: domain 1 of 1	36..313	38/311 (12%) 150/311 (48%)	7.9

**EXAMPLE 33.**

The NOV33 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis			
	SEQ ID NO: 145	1240 bp	
NOV33a, CG59430-01 DNA Sequence	ACCAAGGACCCCAGAGGATGGAGGCCTCTCGGTGGTGGCTGCTGGTCACTGTGCTCAT GGCTGGGGCTCATTGTGTGGCCCTGGTTGACCAAGAAGCTTCTGATCTCATCCATTCT GGCCCCAGGACAGCAGCCCTGGGCCTGCCCTGCCCTGCCACAAAATCTCTGTGAGCA ACATAGACTTTGCGTTCAAGCTCTACAGACAGTTGGCTTTGAACGCCCCCGGGGAGAA CATTCTCTTCTCCCCAGTGAGCATCTCCCTGGCCTTGGCCATGCTTTCTTGGGGGGCC CCAGTGGCCAGCAGGACCCAACTCCTGGAGGGCCTGGGGTTACCCCTCACCGTGGTGC CTGAGGAGGAGATCCAGGAAGGCTTCTGGGATCTGCTGATCAGGCTCCGTGGGCAGGG TCCCCGGCTCCTCCTGACCATGGACCAGCGCAGGTTACAGCGCCTGGGCGCAGGGCC AACCAGAGCCTAGAGGAGGCCAAAAACACATTGACGAATATACAGAGCAGCAGACCC AGGGGAAGCTCGGGGCTTGGGAGAAGGACCTCGGCAGTGAAACCACAGCGGTTCTGGT GAATCACATGCTCCTCAGAGCTGAGTGGATGAAGCCCTTTGACTCACGTGCCACCAGC CCAAAGGAGTTCTTTGTAGATGAGCACAGCGCTGTGTGGGTGCCCATGATGAAGGAGA AGGCCAGCCACCGCTTCTGACGACCGTGAGCTGCAATGCTCTGTGCTGCGGATGGA CCACGCTGGGAACACCACCACCTTCTTCATCTTCCCCAACAGGGGCAAGATGAGGCAG CTGGAAGATGCCCTGCTGCCTGAAACACTGATTAAAGTGGGACAGTCTGCTCAGGCTCC ATTTCCACTTCCCCAAATTTTCCATTCTAGAACCTGCAGACTGGAGATGCTCCTCCC AAAAGTGACTGTGGGTGGAGGCTTCCCTGGGCAGCCTGCACTGAACATTTCTAAAGTA AGTTGGGGATGGTGTGTTTCAAGGGCCTCTCATAAGGCCATGATGACGCTGGATGAGA GGGGCTCTGAAGCTGCTGCAGCCACCAGCATTAGCTCACCCCTGGGCCTCGCCGAGA CCTTGACTTCCCACCACCTCTGGGCACTGAGTTCACTCGGCCCTTCTGCTGATGACT TTCCACACGGAAACAGGAAGCATGCTTTTCTGGAGAAGATTGTAAACCACTGGGAT <u>AACGCCCCCTCAGACATGCTGG</u>		
	ORF Start: ATG at 18	ORF Stop: TAA at 1218	
	SEQ ID NO: 146	400 aa	MW at 44726.0kD
NOV33a, CG59430-01 Protein Sequence	MEASRWLLVTVLMAGAHCVLVDQEASDLIHSGPQDSSPGPALPCHKISVSNIDFAE KLYRQLALNAPGENILFSPVSI SLALAMLSWGPVASRTQLLEGLGFTLTVPPEEIQ EGFWDLLIRLRGQGRLLLTMDQRRFSGLGARANQSLEEAQKHIDEYTEQQTQGKLG WEKDLGSETTAVLVNHMLLRAEWMKPFDSRATSPKEFFVDEHSAVWVPMMEKASHRE LHDRELQCSVL RMDHAGNTTTFIFPNRGKMRQLEDALLPETLIKWDSLLRLDFHFPK FSISRTCRLEMLLPKVTVGGGFPQGPLNISKVSWGVCVRASHKAMMTLDERGSEAF AATSIQLTPGPRPDLDFPPTLGTEFSRPFVMTFHTETGSMLFLEKIVNPLG		

Further analysis of the NOV33a protein yielded the following properties shown in

## 5 Table 33B.

Table 33B. Protein Sequence Properties NOV33a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1700 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)

SignalP analysis:	Likely cleavage site between residues 20 and 21
-------------------	---

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 33C.

Table 33C. Geneseq Results for NOV33a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68434	Amino acid sequence of human serpin protease Zserp11 - Homo sapiens, 366 aa. [WO200138534-A2, 31-MAY-2001]	1..400 1..366	352/400 (88%) 356/400 (89%)	0.0
AAO13910	Human polypeptide SEQ ID NO 27802 - Homo sapiens, 495 aa. [WO200164835-A2, 07-SEP-2001]	7..399 80..493	164/431 (38%) 234/431 (54%)	4e-67
AAG73736	Human colon cancer antigen protein SEQ ID NO:4500 - Homo sapiens, 446 aa. [WO200122920-A2, 05-APR-2001]	7..399 31..444	164/431 (38%) 234/431 (54%)	4e-67
AAY28643	Human serine protease inhibitor from cDNA clone HETDK50 - Homo sapiens, 422 aa. [WO9940183-A1, 12-AUG-1999]	7..399 7..420	164/431 (38%) 234/431 (54%)	4e-67
AAB74691	Human protease and protease inhibitor PPIM-24 - Homo sapiens, 422 aa. [WO200110903-A2, 15-FEB-2001]	7..399 7..420	163/431 (37%) 233/431 (53%)	1e-66

In a BLAST search of public sequence databases, the NOV33a protein was found to  
5 have homology to the proteins shown in the BLASTP data in Table 33D.

Table 33D. Public BLASTP Results for NOV33a				
Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC42686	SEQUENCE 1 FROM PATENT WO0138534 - Homo sapiens (Human), 366 aa.	1..400 1..366	352/400 (88%) 356/400 (89%)	0.0
Q96BZ5				5e-66

	PROTEIN - Homo sapiens (Human), 427 aa.	9..424	235/423 (55%)	
P29622	Kallistatin precursor (Kallikrein inhibitor) (Protease inhibitor 4) - Homo sapiens (Human), 427 aa.	8..398 9..424	161/423 (38%) 235/423 (55%)	1e-65
P05544	Contrapsin-like protease inhibitor 3 precursor (CPI-23) (Serine protease inhibitor 1) (SPI-1) - Rattus norvegicus (Rat), 413 aa.	12..398 9..412	151/412 (36%) 222/412 (53%)	3e-60
S08102	serine proteinase inhibitor 1 - rat, 403 aa.	36..398 22..402	145/388 (37%) 213/388 (54%)	4e-60

PFam analysis predicts that the NOV33a protein contains the domains shown in the Table 33E.

Table 33E. Domain Analysis of NOV33a			
Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin: domain 1 of 3	46..137	48/93 (52%) 74/93 (80%)	8.9e-31
serpin: domain 2 of 3	154..306	68/168 (40%) 105/168 (62%)	6.5e-34
serpin: domain 3 of 3	332..398	31/71 (44%) 53/71 (75%)	1.1e-14

#### EXAMPLE 34.

The NOV34 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis		
	SEQ ID NO: 147	1026 bp
NOV34a, CG59305-01 DNA Sequence	<p>ATGCCACAGCCCAGGGGAGGCCAGCCTGCCTGGCAGCTGACACCCAGCCCTCCCCCA GCTCCCGGATAATGAGCACCCATGTGGCAGGCCTGGGCCTGGACAAGATGAAGCTGGG CAATCCCCAGTCCTTCCTGGACCAGGAGGAGGCAGATGACCAGCAGCTGCTGGAACCA GAGGCGTGGAAGACCTACACCGAGCGCCGCAATGCCCTGCGTGAGTTCCTGACCTCGG ACCTGAGTCCGCACCTGCTCAAGCGCCACCACGCCCGCATGCAGCTGCTGCGTAAGTG CTCCTACTACATCGAGGTCTGCCCAAGCACCTGGCCCTGGGCGACCAGAACCCGCTG GTGCTGCCTAGCGCCTTGTTCAGCTCATCGACCCCTGGAAGTTCCAGCGCATGAAGA AGGTGGGCACAGCTCAGACCAAGATCCAGCTCCTGCTGCTCGGGACCTGTTGGAACA GCTCGACCATGGCCGTGCTGAGCTGGATGCCCTGCTCCGTCGCCAGACCCACGGCCC TTCCTGGCCGACTGGGCGCTGGTGAGCGGCGGCTGGCGGACGTGTCCGCCGTGATGG ACAGCTTCCTGACCATGATGGTGCCGGGGCGGCTACACGTCAAGCACCGCCTGGTGTC TGATGTGAGTGCCACCAAGATCCCGCACATCTGGCTCATGTGAGCACCAAGATGCCT GTCGTGTTTGACCGAAAGGCGTCGGCGGCTCACCAGGACTGGGCCCGGCTGCGCTGGT TCGTACCATCCAGCCAGCCACATCGGAGCAGTATGAGTTGCGCTTCAGGCTGCTGGA CCCGCGACACAGCAGGAGTGCGCCAGTGTGGCGTCATCCCGTGGCTGCCTGCACC</p>	

	TTCGACGTCCGAAACCTGCTGCCAACCGATCCTATAAGTTCACCATCAAGAGGGCCG AGACCTCCACGCTGGTGTACGAGCCCTGGAGGGACAGCCTCACCTGCACACCAAGCC GGAGCCCCTGGAGGGGCCGCCCTCAGCCACTCTGTCTGA		
	ORF Start: ATG at 1	ORF Stop: TGA at 1024	
	SEQ ID NO: 148	341 aa	MW at 38993.7kD
NOV34a, CG59305-01 Protein Sequence	MPQPRGGQPAWQLTPSPPPSSRIMSTHVAGLGLDKMKLGNPQSFLDQEEADDQQLLEP EAWKTYTERRNALREFLTSDLSPHLLKRHHARMQLLRKCSYYIEVLPHKHLALGDQNPL VLPSALFQLIDPWKFQRMKKVGTATQTKIQLLLLDLLEQLDHGRAELDALLRSPDPRP FLADWALVERRLADVSAVMDSFLTMVPGRLHVKHRLVSDVSATKIPIHWMLSTKMP VVFDRKASAAHQDWARLRWFVTIQPATSEQYELRFRLDPRTQQECAQCGVIVAACT FDVRNLLPNRSYKFTIKRAETSTLVYEPWRDSLTLHTKPEPLEGPALSHSV		
	SEQ ID NO: 149	1026 bp	
NOV34b, CG59305-02 DNA Sequence	ATGCCACAGCCCAGGGGAGGCCAGCCTGCCTGGCAGCTGACACCCAGCCCTCCCCCA GCTCCCGGATAATGAGCACCATGTGGCAGGCCTGGGCCTGGACAAGATGAAGCTGGG CAATCCCCAGTCTTCTCTGGACCAGGAGGAGGCAGATGACCAGCAGCTGCTGGAACCA GAGGCGTGGAAGACCTACACGAGCGCGCAATGCCCTGCGTGAGTTCTCTGACCTCGG ACCTGAGTCCGCACCTGCTCAAGCGCCACCACGCCCGCATGCAGCTGCTGCGTAAGTG CTCCTACTACATCGAGGTCTGCCCCAAGCACCTGGCCCTGGGCGACCAAGACCCGCTG GTGCTGCCTAGCGCCTTGTTCCAGCTCATCGACCCCTGGAAGTTCAGCGCATGAAGA AGGTGGGCACAGCTCAGACCAAGATCCAGCTCCTGCTGCTCGGGGACCTGTTGGAACA GCTCGACCATGGCCGTGCTGAGCTGGATGCCCTGCTCCGGTGCAGACCCACGGCC TTCCTGGCCGACTGGGCGCTGGTGGAGCGGCGGCTGGCGGACGTGTGCGCCGTATGG ACAGCTTCTTGACCATGATGGTGCCGGGGCGGCTACACGTCAAGCACCGCCTGGTGTC TGATGTCAGTGCCACCAAGATCCCGCACATCTGGCTCATGCTGAGCACCAAGATGCCT GTCGTGTTTGACCGAAAGGCGTCGGCGGCTCACCAGGACTGGGCCCGGCTGCGCTGGT TCGTACCATCCAGCCAGCCACATCGGAGCAGTATGAGTTGCGCTTCAGGCTGCTGGA CCCGCGGACACAGCAGGAGTGCGCCAGTGTGGCGTCATCCCGTGGCTGCTGCAACC TTCGACGTCCGAAACCTGCTGCCAACCGATCCTATAAGTTCACCATCAAGAGGGCCG AGACCTCCACGCTGGTGTACGAGCCCTGGAGGGACAGCCTCACCTGCACACCAAGCC GGAGCCCCTGGAGGGGCCGCCCTCAGCCACTCTGTCTGA		
	ORF Start: ATG at 1	ORF Stop: TGA at 1024	
	SEQ ID NO: 150	341 aa	MW at 38993.7kD
NOV34b, CG59305-02 Protein Sequence	MPQPRGGQPAWQLTPSPPPSSRIMSTHVAGLGLDKMKLGNPQSFLDQEEADDQQLLEP EAWKTYTERRNALREFLTSDLSPHLLKRHHARMQLLRKCSYYIEVLPHKHLALGDQNPL VLPSALFQLIDPWKFQRMKKVGTATQTKIQLLLLDLLEQLDHGRAELDALLRSPDPRP FLADWALVERRLADVSAVMDSFLTMVPGRLHVKHRLVSDVSATKIPIHWMLSTKMP VVFDRKASAAHQDWARLRWFVTIQPATSEQYELRFRLDPRTQQECAQCGVIVAACT FDVRNLLPNRSYKFTIKRAETSTLVYEPWRDSLTLHTKPEPLEGPALSHSV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 34B.

Table 34B. Comparison of NOV34a against NOV34b and NOV34c.		
Protein Sequence	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV34b	1..341	328/341 (96%)
	1..341	328/341 (96%)

Further analysis of the NOV34a protein yielded the following properties shown in Table 34C.

Table 34C. Protein Sequence Properties NOV34a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.4466 probability located in microbody (peroxisome); 0.2245 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 34D.

Table 34D. Geneseq Results for NOV34a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found				

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34E.

Table 34E. Public BLASTP Results for NOV34a				
Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BVV2	HYPOTHETICAL 36.5 KDA PROTEIN - Homo sapiens (Human), 318 aa.	24..341 1..318	318/318 (100%) 318/318 (100%)	0.0
Q9D9W3	1700026M20RIK PROTEIN - Mus musculus (Mouse), 163 aa.	66..173 2..109	89/108 (82%) 96/108 (88%)	6e-46

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34F.

Table 34F. Domain Analysis of NOV34a			
Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
fn3: domain 1 of 1	231..312	10/87 (11%) 52/87 (60%)	5.9

**EXAMPLE 35.**

The NOV35 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis			
	SEQ ID NO: 151	1610 bp	
NOV35a, CG59547-01 DNA Sequence	CGTCTTCGGGACGCGCCCGCTCTTCGCCTTTTCGCTGCAGTCCGTCGATTCTTTCTCC AGGAAGAAAAATGGCATCCGTTGCAGTTGATCCACAACCGAGTGTGGTGACTCGGGTG GTCAACCTGCCCTTGGTGAGCTCCACGTATGACCTCATGTCTCAGCCTATCTCAGTA CAAAGGACCAGTATCCCTACCTGAAGTCTGTGTGTGAGATGNCAGAGAACGGTGTGAA GACCATCACCTCCGTGGCCATGACCAGTGCTCTGCCCATCATCCAGAAGCTAGAGCCG CAAATTGCAGTTGCCGATACCTATGCCTGTAAGGGGCTAGACAGGATTGAGGAGAGAC TGCCATTCTGAATCAGCCATCAACTCAGATTGTTGCCAATGCCAAAGGCGCTGTGAC TGGGGCAAAGATGCTGTGACGACTACTGTGACTGGGGCCAAGGATTCTGTNGCCAGC ACGATCACAGGGGTGATGGACAAGACCAAAGGGGCAGTGAAGTGGCAGTGTGGAGAAGA CCAAGTCTGTGGTCAGTGGCAGCATTAAACACAGTCTTGGGGAGTCGGATGATGCAGCT CGTGAGCAGTGGCGTAGAAAATGCACTCACCAAATCAGAGCTGTTGGTAGAACAGTAC CTCCCTCTCACTGAGGAAGAACTAGAAAAAGAAAGCAAAAAAGTTGAAGGATTTGATC TGTTTCAGAAGCCAAGTTATTATGTTAGACTGGGATCCCTGTCTACCAAGCTTCACTC CCGTGCCTACCAGCAGGCTCTCAGCAGGGTTAAAGAAGCTAAGCAAAAAAGCCAACAG ACCATTCTCAGCTCCATTCTACTGTTACCTGATTGAATTTGCCAGGAAGAATGTGT ATAGTGCCAATCAGAAAATTCAGGATGCTCAGGATAAGCTCTACCTCTCATGGGTAGA GTGGAAGAGGAGCATTGGATATGATGATACTGATGAGTCCCACTGTGCTGAGCACATT GAGTCACGTACTCTTGCAATTGCCCGCAACCTGACTCAGCAGCTCCAGACCACGTGCC ACACCCTCCTGTCCAACATCCTTTGTGTACCACAGAACATCCCCCATCATTTTTTGCA AAAGGGGTGATGGCAGGCGACATCTACTCAGTGTCCGGAATGCTGCCTCCTTTAA GAAGTGTCTGACAGCCTCCTCACTTCTAGCAAGGGGCAGCTGCAGAAAATGAAGGAAT CTTTAGATGACGTGATGGATTATCTTGTTTACAAAACGCCCTAACTGGCTGGTAGC TCCCTTTTATCCTCAGCTGACTGAGTCTCAGAATGCTCAGGACCAAGGTGCAGAGATC GACAAGAGCAGCCAGGAGACCCAGCGATCTGAGCATAAACTCATTAAACCTGCCCT ATCACTAGTGCATGCTGTGGCCAGACAGATGACACCTTTTGTATGTTGAAATTAAC TGCTAGGCAACCTAAATTGGGAAGCAAGTAGCTAGTATAAAGGCCCTCAATTGTAGT TGTTTCCAGCTGAATTAAGAGCTTTAAAGTTTCTGGCATTAGCAGATGATTTCTGTT ACCTGGTAAGAAAAGAATGATAGGCTTGTGAGAGCCTATAGCCA		
	ORF Start: ATG at 69	ORF Stop: TAA at 1380	
	SEQ ID NO: 152	437 aa	MW at 48148.2kD
NOV35a, CG59547-01 Protein Sequence	MASVAVDPQPSVVTRVNLPLVSSYDLMSAYLSTKDQYPYLKSVCEMXENGVKITIT SVAMTSALPIIQKLEPQIAVDITYACKGLDRIERLPILNQPSQIVANAKGAVTGAK DAVTTTITGAKDSVASTITGMDKTKGAVTGSVEKTSVSVSGSINTVLGSRMMQLVSS GVENALTKSELLVEQYLPLTEEELEKEAKKVEGFDLVQKPSYYVRLGSLSTKLHGRAY QQALSRVKEAKQKSQQTISQLHSTVHLIEFARKNVYSANQKIQDAQDKLYLSWVEWKE SIGYDDTDESHCAEHIESRTLAIARNLTQQLQTTCHTLLSNI LCV PQNI PHHFLQKGV MAGDIYSVFRNAASFKEVSDSLTSSKGQLQMKESLDDVMDYLVYKTLNWLVGPFY PQLTESQNAQDQGAEMDKSSQETQRSEHKTH		

Further analysis of the NOV35a protein yielded the following properties shown in

5 Table 35B.

Table 35B. Protein Sequence Properties NOV35a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)







SignalP analysis:	No Known Signal Sequence Predicted
-------------------	------------------------------------

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 35C.

Table 35C. Geneseq Results for NOV35a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY99534	Human adipocyte-specific differentiation-related protein ADRP - Homo sapiens, 437 aa. [WO200031532-A1, 02-JUN-2000]	1..300 138..437	289/300 (96%) 289/300 (96%)	e-161
AAW53264	Human adipocyte-specific differentiation-related protein - Homo sapiens, 437 aa. [US5739009-A, 14-APR-1998]	1..300 138..437	289/300 (96%) 289/300 (96%)	e-161
AAB58800	Breast and ovarian cancer associated antigen protein sequence SEQ ID 508 - Homo sapiens, 250 aa. [WO200055173-A1, 21-SEP-2000]	51..300 1..250	238/250 (95%) 238/250 (95%)	e-133
AAW06798	Murine p154 - Mus sp, 425 aa. [US5541068-A, 30-JUL-1996]	1..298 138..423	231/298 (77%) 256/298 (85%)	e-125
AAR45151	Sequence of mouse adipocyte polypeptide (ap) p154 - Acomys cahirinus, 425 aa. [US5268295-A, 07-DEC-1993]	1..298 138..423	231/298 (77%) 256/298 (85%)	e-125

In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35D.

Table 35D. Public BLASTP Results for NOV35a				
Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC09025	SEQUENCE 22 FROM PATENT WO0031532 - Homo sapiens (Human), 437 aa.	1..300 138..437	289/300 (96%) 289/300 (96%)	e-160
Q9BSC3				e-160

	RELATED PROTEIN - Homo sapiens (Human), 437 aa.	138..437	289/300 (96%)	
Q99541	Adipophilin (Adipose differentiation-related protein) (ADRP) - Homo sapiens (Human), 437 aa.	1..300 138..437	287/300 (95%) 287/300 (95%)	e-159
Q9TUM6	Adipophilin (Adipose differentiation-related protein) (ADRP) - Bos taurus (Bovine), 450 aa.	1..282 138..419	239/282 (84%) 258/282 (90%)	e-132
Q9MZE5	ADIPOSE DIFFERENTIATION-RELATED PROTEIN - Sus scrofa (Pig), 404 aa (fragment).	1..267 138..404	231/267 (86%) * 246/267 (91%)	e-127

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35E.

Table 35E. Domain Analysis of NOV35a			
Pfam Domain	NOV35a Match Region	Identities/ Similarities for the Matched Region	Expect Value
SPX: domain 1 of 1	29..153	24/347 (7%) 80/347 (23%)	8
perilipin: domain 1 of 1	1..259	166/411 (40%) 247/411 (60%)	7.4e-89

#### EXAMPLE 36.

The NOV36 clone was analyzed, and the nucleotide and predicted polypeptide

5 sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis			
	SEQ ID NO: 153	355 bp	
NOV36a, CG58508-01 DNA Sequence	ACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTCTGTCTCTCCTCGTGCT AGTAGCTGCCTTCTGCTCTCTAGCACTCTCAGCACCAATGGGCTCAGACCCCTCCCAC GCCTGCTGCTTTTCTTACACCGCGAGGAAGCTTCCTCACAACCTTTGTGGTAGATTACT ATGAGACCAGCAGCCTCTGCTCCCAGCCAGCTGTGGTATTCCAAACCAAAAGAGGCAA GCAAGTCTGCGCTGACCCAGTGAGTCTGGGTCCAGGAGTACGTGTATGACCTGGAA CTGAAC <u>TGAGCTGCTCAGAGACAGGACAGTCACGCAGAGCTTCATGGTATTGGTGGCA</u> <u>AAGAGGT</u>		
	ORF Start: ATG at 21	ORF Stop: TGA at 297	
	SEQ ID NO: 154	92 aa	MW at 10146.6kD
NOV36a, CG58508-01 Protein Sequence	MKLCVTVLSLLVLVAAFCSLALSAPMGSDPPTACCFSYTARKLP HNFVVDYYETSSL SQPAVVVFQTKRGKQVCADPSES <del>WV</del> QEYVYDLELN		
	SEQ ID NO: 155	355 bp	

NOV36b, CG58508-02 DNA Sequence	ACCTCTTTGCCACCAATACCATGAAGCTCTGCGTGACTGTCCTGTCTCTGAGCAGCTC AGTTTCAGTTCAGGTCATACACGTACTCCTGGACCCAGGACTCACTGGGGTCAGCGCA GACTTGCTTGCCCTCTTTTGGTTTGGAAATACCACAGCCGGCTGGGAGCAGAGGCTGCTG GTCTCATAGTAATCTACCACAAAGTTGCGAGGAAGCTTCTCGCGGTGTAAGAAAAGC AGCAGGCGGTGGGAGGGTCTGAGCCCATTGGTGCTGAGAGTGCTAGAGAGCAGAAGGC AGCTACTAGCACGAGGAGAGACAGGACAGTCACGCAGAGCTTCATGGTATTGGTGGCA AAGAGGT		
	ORF Start: ATG at 21	ORF Stop: TAG at 297	
	SEQ ID NO: 156	92 aa	MW at 10149.6kD
NOV36b, CG58508-02 Protein Sequence	MKLCVTVLSLLVLVAAFCSAPMSAPMGSDPPTACCFSTARKLPRNFVVDYYETSSLC SQPAVVFQTKRGKQVCADPSESQVQEVYVDLELN		
	SEQ ID NO: 157	219 bp	
NOV36c, 170072532 DNA Sequence	GGATCCGCACCAATGGGCTCAGACCTCCCACCGCCTGCTGCTTTTCTTACACCGCGA GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA GCCAGCTGTGGTATTCCAAACCAAAAGAGGCAAGCAAGTCTGCGCTGACCCAGTGAG TCCTGGGTCCAGGAGTACGTGTATGACCTGGAACCTGAACCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 158	73 aa	MW at 8175.1kD
NOV36c, 170072532 Protein Sequence	GSAPMGSDPPTACCFSTARKLPRNFVVDYYETSSLCSPAVVFQTKRGKQVCADPSE SWVQEVYVDLELNLE		
	SEQ ID NO: 159	219 bp	
NOV36d, 170072551 DNA Sequence	GGATCCGCACCAATGGGCTCAGACCTCCCACCGCCTGCTGCTTTTCTTACACCGCGA GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA GCCAGCTGTGGTATTCCAAACCAAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGTGAA TCCTGGGTCCAGGAGTACGTGTATGACCTGGAACCTGAACCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 160	73 aa	MW at 8205.1kD
NOV36d, 170072551 Protein Sequence	GSAPMGSDPPTACCFSTARKLPRNFVVDYYETSSLCSPAVVFQTKRSKQVCADPSE SWVQEVYVDLELNLE		
	SEQ ID NO: 161	219 bp	
NOV36e, 170072555 DNA Sequence	GGATCCGCACCAATGGGCTCAGACCTCCCACCGCTTGCTGCTTTTCTTACACCGCGA GGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGCTCCCA GCCAGCTGTGGTATTCCAAACCAAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGTGAA TCCTGGGTCCAGGAGTACGTGTATGACCTGGAACCTGAACCTCGAG		
	ORF Start: GGA at 1	ORF Stop:	
	SEQ ID NO: 162	73 aa	MW at 8205.1kD
NOV36e, 170072555 Protein Sequence	GSAPMGSDPPTACCFSTARKLPRNFVVDYYETSSLCSPAVVFQTKRSKQVCADPSE SWVQEVYVDLELNLE		
	SEQ ID NO: 163	301 bp	
NOV36f, CG58508-03 DNA Sequence	CAGCCTCACCTCTGAGAAAACCTCTTTCCACCAATACCATGAAGCTCTGCGTGACTG TCCTGTCTCTCCTCATGCTAGTAGCTGCCTTCTGCTCTCCAGCGCTCTCAGCCAGCTG TGGTATTCCAAACCAAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGTGAATCCTGGGT CCAGGAGTACGTGTATGACCTGGAACCTGAACCTGAGCTGCTCAGAGACAGGAAGTCTTC		

	AGGGAAGGTCACCTGAGCCCGGATGCTTCTCCATGAGACACATCTCCTCCATACTCAG GACTCCTCTCA		
	ORF Start: ATG at 40	ORF Stop: TGA at 154	
	SEQ ID NO: 164	38 aa	MW at 3940.8kD
NOV36f, CG58508-03 Protein Sequence	MKLCVTVLSLLMLVA AFCSPALSASCGIPNQKKQASLC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 36B.

Table 36B. Comparison of NOV36a against NOV36b through NOV36f.		
Protein Sequence	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV36b	15..92	76/78 (97%)
	15..92	76/78 (97%)
NOV36c	23..92	69/70 (98%)
	2..71	69/70 (98%)
NOV36d	23..92	68/70 (97%)
	2..71	68/70 (97%)
NOV36e	23..92	68/70 (97%)
	2..71	68/70 (97%)
NOV36f	1..27	23/27 (85%)
	1..27	24/27 (88%)

Further analysis of the NOV36a protein yielded the following properties shown in Table 36C.

Table 36C. Protein Sequence Properties NOV36a	
PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

- 5 A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 36D.

Table 36D. Geneseq Results for NOV36a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/	Identities/ Similarities	Expect Value

		Residues	Matched Region	
AAR36770	MIP-1beta - Homo sapiens, 92 aa. [WO9309799-A, 27-MAY-1993]	1..92 1..92	89/92 (96%) 90/92 (97%)	9e-48
AAB15789	Human chemokine MIP1beta SEQ ID NO: 20 - Homo sapiens, 92 aa. [WO200042071-A2, 20-JUL-2000]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW82717	Human Act-2 protein - Homo sapiens, 92 aa. [WO9854326-A1, 03-DEC- 1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW76225	Human chemokine MIP-1beta domain protein fragment - Homo sapiens, 92 aa. [WO9838212-A2, 03-SEP-1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47
AAW76223	Human chemokine MIP-1beta domain protein from clone MPB-X - Homo sapiens, 331 aa. [WO9838212-A2, 03- SEP-1998]	1..92 1..92	88/92 (95%) 89/92 (96%)	6e-47

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36E.

Table 36E. Public BLASTP Results for NOV36a				
Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P13236	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (T-cell activation protein 2) (ACT-2) (PAT 744) (H400) (SIS-gamma) (Lymphocyte activation gene-1 protein) (LAG-1) (HC21) (G-26 T lymphocyte-secreted protein) - Homo sapiens (Human), 92 aa.	1..92 1..92	88/92 (95%) 89/92 (96%)	2e-46
P46632	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (Immune activation protein 2) (ACT-2) - Oryctolagus cuniculus (Rabbit), 92 aa.	1..92 1..92	75/92 (81%) 84/92 (90%)	7e-39
P50230	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) - Rattus norvegicus (Rat), 92 aa.	1..92 1..92	71/92 (77%) 81/92 (87%)	3e-38

P14097	Small inducible cytokine A4 precursor (Macrophage inflammatory protein 1-beta) (MIP-1-beta) (H400 protein) (SIS-gamma) (ACT2) - Mus musculus (Mouse), 92 aa.	1..92 1..92	69/92 (75%) 82/92 (89%)	1e-36
CAA01323	HUMAN ACT-2 SYNTHETIC GENE PROTEIN - synthetic construct, 74 aa (fragment).	19..92 1..74	69/74 (93%) 69/74 (93%)	2e-35

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36F.

Table 36F. Domain Analysis of NOV36a			
Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8: domain 1 of 1	24..89	25/70 (36%) 60/70 (86%)	2.6e-32

**EXAMPLE 37.**

The NOV37 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis		
	SEQ ID NO: 165	5285 bp
NOV37a, CG59819-01 DNA Sequence	GGCCGGGGAGGGGGCCGGACCGCGCGCGACCGGTGCGCGCCGCTGGGGCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCGG CGCACGGCCGGCTGCGGAGGATCACCTACGTGGTGACCCGGGCCCGGCTGGCAGC CGGCGCCTTGCCCTGAGCGGGCCCCCGCGTTTCGCGGACATTCAACGTGCGCTCAAC GCCAGGTACAGCCGAGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCCTCCCCGGGG TCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGGCGGCCCTGCAGGGGCT CAGACCGCCGCCCGCGCGCGCGGAGCCTGCGCGTCCCGCGGTCCCGGCGGGCAG CTCCACCCCAATCCCGGCGGCCACCCGGCAGCCGCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCCAAGGTGCGCGAGGAGACCCAGAGCGGCGGAGGCTCTAGGCTGCA GGTTCACCAGAAGCAGCAGCTGCAGGGGGTCAATGTCTGTGGAGGGCGGTGCTGTCAT GGCTGGAGTAAGGCCCCCTGGCTCCAGAGGTGCACCAACCTAGCTGTGTTCCGCCAT GTCAGAAATGGAGGGATGTGTCTCCGGCCACAACCTGTGTGTGTAAACCAGGGACCAA GGGCAAGCCTGTGAAACAATAGCTGCCCAGGACACCTCGTCACCAGTCTTTGGAGGG CAGAGTCTGGGGCTGCTTCTCGTGGGGCCCTCCTGAGCAAGCAGCAAGCATACTT CATCTAAGAAGGCAGACACTTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCACTGGGACTCCCCAGCAGATACATTCTCAAGTGACTCCT CTTTCTTCCAGAGCGTGGTTATTACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCAGAAGGGGATTTCAAGGAAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATATGTGCAGGATCAAGTTGCGGCACCTTTTCAGCTGAGTAACCACACT GGCCGCATCAAGGTGGTCTTACTCCGAGCATCTGTAAAGTGACCTGCACCAAGGGCA GCTGTCAGAACAGCTGTGAGAAGGGGAACACCACCACTCTCATTAGTGAGAATGGTCA TGCTGCGGACACCCTGACGGCCACGAACTTCCGAGTGGTAATTTGCCATCTTCCATGT ATGAATGGTGGCCAGTGACGTTCAAGGGACAAATGTGAGTCCCTCCAAATTTACAG GAAAACTTTGTGAGATCCAGTCCATGGTGCCAGCGTGCCATAACTTTATCAGCATTC CCAGCAGCCAGGCAAGGCGTTGGGGACGCATGTATCCATTCACACATACCTTGCCCT CTGACCGTGACTAGCCAGCAAGGAGTCAAAGTGAAATTTCTCCTAACATAGTCAATA TCCATGTGAAACATCCTCCTGAAGCTCCGTCCAGATACATCAGTTTCAAGAATTGA	



TGGCCCAACAGGCCAGAAGACAAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC  
 CAAGGGCTTCTGTCCAGAAGACCCAGACCATACATTCCACATACTCCACCAGCAGG  
 TCATTCTCTACGTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA  
 AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC  
 TGTGGAACTGTGGGTACCTCTGGGGCTTTAACAATGCCAGAAATGCCCAAGAAAC  
 CATCTTATCATGGATACAACCAATGATGGAATGCCTACCGGGTTATAAGCGGGTTAA  
 CAACACCTTTTGCCAAGATATTAATGAATGTCAGCTACAAGGTGTATGCCCTAATGGT  
 GAGTGTGTAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG  
 ATCCTACCTTTTCAAGTTGTGTTCTGATCCCCCTGTGATCTCGGAAGAGAAAGGGCC  
 CTGTTACCGACTTGTCAAGTCTGGAAGACAGTGTATGTACCTCTGTCTGTTACCTC  
 ACCAAGCAGCTCTGCTGTGTAGTGTGGCAAGGCCTGGGGCCACACTGTGAGAAAT  
 GTCCCTTCCAGGCACAGCTGCTTTAAGGAAATCTGTCTGGTGAATGGGTTATAC  
 GGTTTCTGGCGTTTCATAGACGCAGGCCAATCCATCACCATGTAGGTTAAGGACCTGTA  
 TTTGTCAAGCCAAAGAACTCAACCTGTTGCTAAAAGTACTCATCTCCACCTCTCC  
 CAGCCAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCCGGGAACACGGGCCAGGAGT  
 GGCGGAGCCAGAAAGTGGCAACTGCACCCCTGAAAAGGAAATACCTTCATTGGATCAA  
 GAGAAACCAAACTTGAGCCTGGTCAACCCAGCTGTCTCCAGGCATTTCCGCTATTTC  
 ATCTGCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACCTCTGTGCTCTGTTGA  
 AGTAGCTCCTGAAGCTTCTACGTCTAGTGCCAGCCAAGTGATTGCTCTACTCAAGTG  
 ACAGAAATCAATGAATGTACTGTGAACCTGATATCTGTGGAGCAGGACACTGCATTA  
 ACCTACAGTGAGATATACCTGTATATGCTACGAGGGCTACAGGTTCAAGTGAACAACA  
 GAGGAAATGTGTATATTTGATGAGTGTACTCAGGTCCAACACCTCTGCTCCCAGGGC  
 CGCTGTGAAAACACCGAGGGAAGTTTCTTGTGCATTTGCCCAGCAGGATTATGCGCCA  
 GTGAGGAGGGTACTAAGTGCATAGATGTTGACGAATGCCGAGGCCGACGCTCTGTGG  
 GGAGGGGCACTGTGTCAATACTGTGGGGCCTTCCGGTGTGAATACTGTGACAGCGGG  
 TACCGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAATGTTTGAATCCAAGCA  
 CTTGTCCAGATGAGCAGTGTGTGAATCTCCTGGATCTTACCAGTGCCTTCCCTGCAC  
 AGAAGGATCCGAGGCTGGAATGGACAGTGCCTTGATGTGGACGAGTGCCTGGAACCA  
 AACGTCTGCGCAATGGTGATTGTTTCCAACCTTGAAGGCTCTACATGTGTTTCATGCC  
 ACAAGGCTATACCCGACTCCGGACCACAAGCACTGTAGAGATATTGATGAATGTCA  
 GCAAGGGAATCTATGTGTAAACGGGCAGTGCAAAAATACCGAGGGCTCCTTCAGGTGC  
 ACCTGTGGACAGGGGTACAGCTGTGCGCAGCTAAAGACCAGTGTGAAGACATTGATG  
 AATGCCAGCACCGTCATCTCTGTGCTCATGGGCAGTGCAGGAACACTGAGGGCTCTTT  
 TCAATGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGAGACCAGTGTGAAGAT  
 ATCAATGAATGCTTGGAGGACAAGAGTGTGTTGCCAGAGAGGAGCTCATTAATATCTG  
 CAGGGTCTATGATTGTACTTGTCCGGATGGATTTAGCTAGATGACAATAAACATG  
 TCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCGCAAGGGGAGTGCCTA  
 AACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTTCTCAATCTCTGCAGATG  
 GCCGTACGTGTGAAGATATTGATGAATGTGTAAACAACACTGTTTGTGACAGTCAAGG  
 GTTTTGTGACAATACAGTGGCTCCTTCCGCTGCCTCTGTATACAGGCTTTTCAAGCC  
 CCACAGGATGGGCAAGGGTGTGTGGATGTGAATGAATGTGAAGTGTGACTGCTGAGTGGGTGT  
 GTGGTGAAGCCTTCTGTGAAAACGTGGAAGGGTCTTCTGTGCGTGTGTGCTGATGA  
 AAACCAAGAGTACAGCCCATGACTGGGCAGTGCCTGCTCCCGGACCTCCACAGATTTA  
 GATGTAGATGTAGATCAACCCAAAGAAAGAAAGAAAGAAATGCTACTATAATCTCAATG  
 ACGCCAGTCTCTGTGATAATGTGTTGGCCCCCAATGTACGAAACAAGAATGTGCTGTG  
 TACATCAGGCGTGGGATGGGAGATAACTGCGAAATCTTCCCTGCCCGGTCTTGGGA  
 ACTGCTGAGTTCACTGAAATGTGTCCCAAAGGGAAGGTTTTGTGCTGCTGGAGAAT  
 CATCTTCTGAAGCTGGTGGTGAAGTATAAAGATGCAGATGAATGCCTACTTTTGG  
 ACAAGAAATCTGCAAAATGGTTTCTGTTTGAACACTCGGCTGGGTATGAATGCTAC  
 TGTAAGCAAGGGACGTAATGATCTGTGAACTGCAGTGCTTTGATATGGATGAAT  
 GTCAAGACCCAGTAGTTGTATTGATGGCCAGTGTGTTAATACAGAGGGCTCTTACAA  
 CTGCTTCTGTACTACCCCATGGTCTGGATGCGTCAGAAAAAGATGTATACGACCG  
 GCTGAGTCAAACGAACAATAAGAAACTGATGTCTACCAAGATTTGTGCTGGGAAC  
 ATCTGAGTGATGAATACGTGTGTAGCCGGCCTCTGTGGGCAAGCAGACAACGTACAC  
 TGAGTGTGCTGTCTGTATGGAGAGGCTGGGCGATGCAGTGTGCCCTCTGCCCCCTG  
 AAGGATTCAGATGACTATGCTCAGCTGTGTAACATCCCGGTGCGGACGCCCGGACGC  
 CATATGGACGGGACGCTTGGTTGACTTCAGTGAACAGTATACTCCAGAAGCCGATCC  
 CTACTTCATCCAAGACCGTTTTCTAAATAGCTTTGAGGAGTTACAGGCTGAGGAATGC  
 GGCATCCTCAATGGATGTGAAAATGGTCTGCTGTGTGAGGGTCCAGGAAGGTTACACCT  
 GCGATTGCTTTGATGGGTATCACTTGGATACTGCCAAGATGACCTGTTTCGATGTAAA  
 TGAATGCGATGAGTTGAACAACCGGATGTCTCTGCAAGAATGCCAAGTGCATTAAC

	ACCGATGGTTCTCTACAAGTGTTTGTGTCTGCCAGGCTACGTGCCTTCTGACAAGCCAA ACTACTGCACTCCGTTGAATACCGCCTTGAATTAGAGAAAGACAGTGACCTGGAGTG AAACAGAATCTACATAACCTAAGCCCATATACTCTGCACGTGTGTAAGGAAAAGGGAG AAATGTA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5219	
	SEQ ID NO: 166	1721 aa	MW at 186900.6kD
NOV37a, CG59819-01 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRITYVVHPGPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVVRSKVPQETQSGGSRLLQVHQKQLQGVNVCGGRCC HGWSKAPGSQRCTKPSCVPPCQNGGMCLRPQLCVCKPGTKGKACETIAAQDTSSPVFG GQSPGAASSWGPPEQAAKHTSSKKADTLPRVSPVAQMTLTLPKPKPSVGLPQQIHSQVI PLSSQSVVIHHGQTQEYVLPKPYFPAQKGISGEQSTEGSFPLRYVQDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNNTTLISENGHAADTLTATNFRVVICHL CMNGGQCSSRDCKQCPCPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVKFPPNINIVHKHPPEASVQIHQVSRIDGPTGQKTKEAQPGSQVS YQGLPVQKTQTIHSTYSHQQVI PHVYPVAAKTQLGRCFQETIGSQCGKALPGLSKQED CCGTVGTSWGFNKCQCKPKPSYHGYNQMMCECLPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVPDPPVISEEKGPCYRLVSSGRQCMYPLSVH LTKQLCCCSVGKAWGPHCEKCPPLPGTAAFKEICPGMGYTVSGVHRRRPIHHVVGKGP VFVKPKNTQPVAKSTHPPPLPAKEEPVEALTFSREHGPVGEVATAPPEKEIPLSD QEKTKLEPGQPQLSPGISAIHLHPQFPVIEKTSPPVPVEVAPEASTSSASQVIAPTQ VTEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCVYIDECTQVQHLCSQ GRCENTEGSFCLICPAGFMASEEGTNCIDVDECLRPDVCGBGHCVNTVGAFRCEYCD GYRMTQRGRCEIDDECLNPSTCPDEQCVNSPGSYQCVPTCEGFRGWNQCLDVECLE PNVCANGDCSNLEGSYMSCCHKGYTRTPDHKHCRIDECQQGNLVCNGQCKNTEGSFR CTCGQGYQLSAAKDQCEDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCE DINECLEDKSVCQRGDCINTAGSYDCTCPDGFQLDDNKTCDINECEHPGLCGPQGE LNTGESHFCVCQCGFSISADGRTCEDIDECVNNTVCDSHGFCDNTAGSFRCLCYQGFQ APQDGGQCVDVNECELLSGVCGEAFCEVNEGSFLVCADENQEYSPMTGQCRSRSTD LDVDVDQPKKEKKECYNLDASLCDNVLPNVTKECCCTSGVGWGDNCIEIFPCPVL GTAEFTMCPKKGKGFVPAGESSEAGGENYKDADECLLFGQEI CKNGFCLNTRPGYEC YCKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGSYNCFCTHPMVLDASEKRCIR PAESNEQIEETDVYQDLWEHLSDEYVCSRPLVGKQTTYTECCCLYGEAWAMQCALCP LKSDDDYAQLCNI PVTGRRQPYGRDALVDFSEQYTPEADPYFIQDRFLNSFEELQAE CGILNGCENGRCVRVQEGYTCDFDGYHLD TAKMTCFDVNECDELNNRMSLCKNAKCI NTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKDSLE		
	SEQ ID NO: 167	5126 bp	
NOV37b, CG59819-02 DNA Sequence	GGCCGGGGGAGGGGGCCGGACCGCGCGCACCGGTGCGGCCCCGCTGGGGCCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCGG CGCACGGCCGGCTGCGGAGGATCACTACGTGGTGACCCGGGCCCCGGCTTGGCAGC CGGCGCCTTGCCCTGAGCGGGCCCCCGGTTGCGGACATTCAACGTGCGCTCAAC GCCAGGTACAGCCGACGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCTCCCCGGGG TCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGGCGCGGCCTGCAGGGGCT CAGACCGCCGCGCGCGCGCGCGGAGCCTGCGCGTCCCGCGGTCCCGCGCGGGCAG CTCCACCCCAATCCCGGCGGCCACCCGGCAGCCGCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCCAAGGTGCCGAGGAGACCCAGAGCGGCGGAGGCTCTAGGCTGCA GGTTCACGAGAAGCAGCAGCTGCAGGGGTCAATGTCTGTGGAGGGCGGTGCTGTCTAT GGCTGGAGTAAGGCCCCCTGGCTCCAGAGGTGCACCAAACCTAGCTGTGTTCCGCCAT GTCAGAAATGGAGGGATGTGTCTCCGGCCACAACCTCTGTGTGTGTAACCAGGGACCAA GGGCAAGCCTGTGAAACAATAGCTGCCAGGACACCTCGTACCAAGTCTTTGGAGGG CAGAGTCTCTGGGGCTGCTTCTCGTGGGGCCCTCCTGAGCAAGCAGCAAGCATACTT CATCTAAGAAGGCAGACACTCTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCACTGGGACTCCCCAGCAGATACATTCTCAAGTGACTCCT CTTTCTTCCAGAGCGTGGTTATTACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCAGAAGGGGATTTCAAGGAGAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATATGTGAGGATCAAGTTGCGGCACCTTTTCAGTGTAGTAACCACT GGCCGCATCAAGGTGGTCTTTACTCCGAGCATCTGTAAGGTGACCTGCACCAAGGCA GCTGTCAAGACAGCTGTGAGAAGGGGAACACCACCACTCTCATTAGTGAGAATGGTCA		

TGCTGCCGACACCCTGACGGCCACGAACTTCCGAGTGGTAATTTGCCATCTTCCATGT  
 ATGAATGGTGGCCAGTGCAGTTCAAGGGACAAATGTCAGTGCCTCCAAATTTACAG  
 GAAAACTTTGTAGATCCAGTCCATGGTGCCAGCGTGCCTAAACTTTATCAGCATT  
 CCAGCAGCCAGGCCAAGGCGTTGGGGACGCATGTCTATCCATTCAACACATACCTTGCT  
 CTGACCGTGACTAGCCAGCAGGAGTCAAAGTGAATTTCTCTCAACATAGTCAATA  
 TCCATGTGAAACATCTCTGAAGCTTCCGTCCAGATACATCAGGTTTCAAGAATTGA  
 TGGCCCAACAGGCCAGAAGACAAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC  
 CAAGGGCTTCTGTCCAGAAGACCCAGACCATAACATCCACATACTCCACCAGCAGG  
 TCATTCTCACGTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA  
 AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC  
 TGTGGAACCTGTGGGTACCTCTGGGGCTTTAACAAATGCCAGAAATGCCCAAGAAAC  
 CATCTTATCATGGATACAACCAATGATGGAATGCCTACCGGTTATAAGCGGGTTAA  
 CAACACCTTTTGCCAAGATATTAATGAATGTCAGCTACAAGGTGTATGCCCTAATGGT  
 GAGTGTTTGAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG  
 ATCCTACCTTTTCAAGTTGTCTCTGATCCCCCTGTGATCTCGGAAGAGAAAGGGCC  
 CTGTTACCGACTTGTGAGTCTGGAAGACAGTGTATGCACCTCTGTCTGTTCACTC  
 ACCAAGCAGCTCTGCTGTTGTAGTGTGGCAAGGCTGGGGCCACACTGTGAGAAAT  
 GTCCCTTCCAGGCACAGCCAAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCGGGA  
 ACACGGGCCAGGAGTGGCGGAGCCAGAAGTGGCAACTGCACCCCTGAAAGGAAATA  
 CCTTCATGGATCAAGAGAAAACCAACTTGAGCTGGTCAACCCAGCTGTCTCCAG  
 GCATTTCCGCTATTCTGTCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACC  
 TCCTGTGCTGTTGAAGTAGCTCTGAAGCTTCTACGTCTAGTGCCAGCCAGTGAAT  
 GCTCCTACTCAAGTGACAGAAATCAATGAATGTACTGTGAACCTGATATCTGTGGAG  
 CAGGACACTGCATTAACCTACCAGTGAGATATACCTGTATATGTACGAGGGCTACAG  
 GTTCAGTGAACAACAGAGGAAATGTGTGTATATGATGAGTGTACTCAGGTCCAACAC  
 CTCTGCTCCAGGGCCGCTGTGAAAACACCGAGGGAAGTTTCTTGTGCAATTTGCCAG  
 CAGGATTTATGGCCAGTGAGGAGGGTACTAAGTGCATAGATGTTGACGAATGCCTGAG  
 GCCGGACGCTCTGTGGGAGGGGCACTGTGTCAATACTGTGGGGCCCTTCCGGTGTGAA  
 TACTGTGACAGCGGGTACCGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAAT  
 GTTTGAATCCAAGCACTTGTCCAGATGAGCAGTGTGTGAATTTCTCTGGATCTTACCA  
 GTGCGTTCCCTGCACAGAAGGATTCCGAGGCTGGAATGGACAGTGCCTTGATGTGGAC  
 GAGTGCCTGGAACCAACGCTCTGCGCAAATGGTGATTGTTCCAACCTTGAAGGCTCCT  
 ACATGTGTTTATGCCACAAAGGCTATACCCGACTCCGACCACAAGCACTGTAGAGA  
 TATTGATGAATGTCAGCAAGGGAATCTATGTGTAACCGGGCAGTGCATAAATACCGAG  
 GGCTCCTTCAGGTGCACCTGTGGACAGGGGTACCAGCTGTCCGCAGCTAAAGACCAGT  
 GTGAAGACATTGATGAATGCCAGCACCGTCATCTCTGTGCTCATGGGCAGTGCAGGAA  
 CACTGAGGGCTCTTTTCAATGTGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGA  
 GACCACTGTGAAGATATCAATGAATGCTTGGAGGACAAGAGTGTGTCAGAGAGGAG  
 ACTGCATTAATACTGCAGGGTCTATGATTGTACTTGTCCGATGGATTCTCAGGTAGA  
 TGACAATAAAACATGTCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCG  
 CAAGGGGAGTGCCTAAACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTCT  
 CAATCTCTGCAGATGGCCGTACGTGTGAAGATATTGATGAATGTGTAACAACACTGT  
 TTGTGACAGTACCGGGTTTTGTGACAATACAGCTGGCTCCTTCCGCTGCCTCTGTTAT  
 CAGGGCTTTCAAGCCCCACAGGATGGGCAAGGGTGTGTGGATGTGAATGAATGTGAAC  
 TGCTCAGTGGGGTGTGTGGTGAAGCCTTCTGTGAAAACGTGGAAGGGTCTTCTGTG  
 CGTGTGTGCTGATGAAAACCAAGAGTACAGCCCCATGACTGGGCAGTGGCGCTCCCGG  
 ACCTCCACAGATTTAGATGTAGATGTAGATCAACCCAAAGAAGAAAAGAAAGATGCT  
 ACTATAATCTCAATGACGCCAGTCTCTGTGATAATGTGTTGGCCCCAATGTACAGAA  
 ACAAGAATGCTGCTGTACATCAGGCGTGGGATGGGAGATAACTGCGAAATCTTCCCC  
 TGCCCGGCTTTGGGAACTGCTGAGTTCACTGAAATGTGTCCCAAAGGGAAAGGTTTG  
 TGCCTGCTGGAGAATCATCTTCTGAAGCTGGTGGTGAGAACTATAAAGATGCAGATGA  
 ATGCCTACTTTTTGGACAAGAAATCTGCAAAAATGGTTTCTGTTTGAACACTCGGCCT  
 GGGTATGAATGCTACTGTAAGCAAGGGACGTAATGATCCTGTGAACTGCAGTGCT  
 TTGATATGGATGAATGTCAAGACCCAGTAGTTGTATTGATGGCCAGTGTGTTAATAC  
 AGAGGGCTCTTACAACTGCTTGTACTCAACCCATGGTCCCTGAGTGCCTGCAGAAAAA  
 AGATGTATACGACCGGCTGAGTCAAACGAACAAATAGAAGAACTGATGTCTACCAAG  
 ATTTGTGCTGGGAACATCTGAGTGATGAATACGTGTGTAGCCGGCCTCTGTGGGCAA  
 GCAGACAACGTACACTGAGTCTGCTGTCTGTATGGAGAGGCTGGGCGATGCAGTGT  
 GCCCTCTGCCCCCTGAAGGATTGAGATGACTATGCTCAGCTGTGTAACATCCCCGTGA  
 CGGGACGCCGCGCAGCCATATGGACGGGACGCTTGGTTGACTTCAGTGAACAGTATAC  
 TCCAGAAGCCGATCCCTACTTCATCCAAGACCGTTTCTAAATAGCTTTGAGGAGTTA

	CAGGCTGAGGAATGCGGCATCCTCAATGGATGTGAAAATGGTCGCTGTGTGAGGGTCC AGGAAGGTTACACCTGCGATTGCTTTGATGGGTATCACTTGGATACTGCCAAGATGAC CTGTTTCGATGTAAATGAATGCGATGAGTTGAACAACCGGATGTCTCTCTGCAAGAAT GCCAAGTGCATTAACACCGATGGTTCCTACAAGTGTGTTGTGTCTGCCAGGCTACGTGC CTTCTGACAAGCCAACTACTGCACTCCGTTGAATACCGCCTTGAATTTAGAGAAAGA CAGTGACCTGGAGTGAACAGAACTACATAACCTAAGCCCATATACTCTGCACCTGTG TAAAGGAAAAGGGAGAAATGTA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5060	
	SEQ ID NO: 168	1668 aa	MW at 181174.9kD
NOV37b, CG59819-02 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRITYVVHGPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVVRSKVPQETQSGGSRQLVHQKQLQGVNVCGRCC HGWSKAPGSQRCTKPSCVPPCQNGMCLRPQLCVCKPGTKGACETIAAQDTSPPVFG GQSPGAASSWGPPPEQAAKHTSSKKADTLPRVSPVAQMTLLTKPKPSVGLPQQIHSQVT PLSSQSVVIHHGQTQEYVVLKPKYFPAQKGISGEQSTEGSFPLRYVDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNNTTLISENGHAADTLTATNFRVVIHCLP CMNGGQCSSRDKQCPCPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVKFPPNIVNIHVKHPPEASVQIHQVSRIDGPTGQKTKEAQPQGSQVS YQGLPVQKTQTIHSTYSHQQVIPHVYPVAAKTQLGRCFQETIGSCQCKALPGLSKQED CCGTVGTSWGFNKCQKCPKPKPSYHGYNQMMCECLPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVDPDPVISEEKGPCYRLVSSGRQCMHPLSVH LTKQLCCSSVGKAWGPHCEKCPPLPGTAKEEPVEALTFSREHGPVABPEVATAPPEKE IPSLDQEKTKLEPGQPQLSPGISAIHLHPQFPVVIKTSPPVPVEVAPEASTSSASQV IAPTQVTEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCVYIDECTQVQ HLCSSQRCENTEGSFCLICPAGFMASEEGTNCIDVDECLRPDVCGEGHCNVTVGAFC EYCDSGYRMTQRGRCEIDDECLNPSTCPDEQCVNSPGSYQCVPCTEGFRGWNGQCLDV DECLPNVCANGDCSNLEGSYMCSCCHKGYTRTPDHKHCRIDECQQGNLCVNGQCKNT EGSFRCCTCGGQYQLSAAKDQCBIDIECQHRHLCAHGQCRNTEGSFQCVCDDQGYRAGSL GDHCEDINECLEDKSVCQRGDCINTAGSYDCTCPDGFQLDDNKTQDINECEHPGLCG PQGECLNTEGSFHCVCQQGFISISADGRTCEDIDECVNNTVCDSHGFCNDTAGSFRLCLC YQGFQAPQDQGQGVVDNECELLSGVCGEAFCEVNEGSFCLVCADENQEYSPMTGQCRS RTSTDLDVDVDQPKKEKKECYYNLDASLCDNVLPNVTKQECCTSGVGWGDNCBIF PCPVLGTAEFTMCPKKGFPVAGESSEAGGENYKDADECLFGQEIKNNGFCMLNTR PGYECYKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGSYCNCFTHPMVLDASE KRCIRPAESNEQIEETDVYQDLWEHLSDHEYVCSRPLVGKQTTYTECCCLYGEAWAMQ CALCPLKDSDDYAQLCNIPTVGRRQPYGRDALVDFSEQYTPADPYFIQDRFLNSFEE LQAEECGILNGCENGRCVRVQEGYTCDFDGYHLDTAKMTCFDVNECELDNNRMSLCK NAKCINTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKDSLE		
	SEQ ID NO: 169	6074 bp	
NOV37c, CG59819-03 DNA Sequence	GGCCGGGGGAGGGGGCCGGACCGCGCGCGACCGGTGCGCGCCCGCTGGGGCCCGCGATG GCGGGGGCCTGGCTCAGGTGGGGGCTCCTGCTCTGGGCAGGGCTCCTCGCGTCTCGG CGCACGGCCGGCTGCGGAGGATCACCTACGTGGTGACCGCGGGCCCCGGCCTGGCAGC CGGCGCCTTGCCCCCTGAGCGGGCCCCCGCGTTGCGCGACATTCAACGTGCGGCTCAAC GCCAGGTACAGCCGAGCTCGGCGGCTGCCGGCGCCCCAGCCGTGCCTCCCCGGGG TCCCCCTCGGAGAGGACCCGGCGCACGAGCAAGCCGGGCGCGCGGCCCTGCAGGGGCT CAGACCGCGCGCGCCGCGCGCGCGGAGCCTGCGCGTCCCGCGGTCCCCGGCGGGCAG CTCACCCCAATCCCGCGGCCACCCGGCAGCGCCCCGTTACCAAACAAGGCAGGC AAGTTGTGCGCTCAAGGTGCCGAGGAGACCCAGAGCGCGGAGGCTCTAGGCTGCA GGTTCACCAAGAAGCAGCAGCTGCAGGGGGTCAATGTCTGTGGAGGGCGGTGCTGTGAT GGCTGGAGTAAGGCCCCGTGGCTCCCAGAGGTGCACCAAACCTAGCTGTGTTCCGCCAT GTCAGAATGGAGGGATGTGTCTCCGGCCACAACCTGTGTGTGTAAACCAGGGACCAA GGGCAAAGCCTGTGAAACAATAGCTGCCAGGACACCTCGTCACCAAGTCTTTGGAGGG CAGAGTCTGGGGCTGCTTCTCCTGCGGGGCCCTCCTGAGCAAGCAGCAAAGCATACTT CATCTAAGAAGGCAGACACTTACCAAGAGTCAGCCCTGTGGCCAGATGACCTTAAC CCTCAAGCCGAAGCCTTCACTGGGACTCCCCCAGCAGATACATTCTCAAGTGACTCCT CTTTCTTCCAGAGCGTGGTTATTACCATGGCCAGACCCAGGAATACGTGCTCAAGC CCAAGTACTTTCCAGCCAGAAGGGGATTTACAGGAACAGTCCACTGAAGGTTCTTT CCCTTTAAGATGTGTCAGGATCAAGTTGCGGCACCTTTTCACTGAGTAACCACT		

GGCCGCATCAAGGTGGTCTTTACTCCGAGCATCTGTAAAGTGACCTGCACCAAGGGCA  
 GCTGTGAGAACAGCTGTGAGAAGGGGAACACCACCCTCTCATTAGTGAGAATGGTCA  
 TGCTGCCGACACCCTGACGGCCACGAACTTCCGAGTGGTAATTTGCCATCTTCCATGT  
 ATGAATGGTGGCCAGTGCAGTTCAAGGGACAAATGTGAGTGCCTCCAAATTTACAG  
 GAAAACTTTGTGAGATCCAGTCCATGGTGCCAGCGTGCCTAAACTTTATCAGCATTC  
 CCAGCAGCCAGGCAAGGCGTTGGGGACGCATGTCATCCATTCAACACATACCTTGCCCT  
 CTGACCGTGACTAGCCAGCAAGGAGTCAAAGTGAAATTTCTCTTAACATAGTCAATA  
 TCCATGTGAAACATCTCTGAAAGCTTCCGTCCAGATACATCAGGTTTCAAGAATTGA  
 TGGCCCAACAGGCCAGAAGACAAAGAAGCTCAACCAGGCCAATCCCAAGTCTCGTAC  
 CAAGGGCTTCTGTCCAGAAGACCCAGACCATACATTCCACATACTCCACCAGCAGG  
 TCATTCTCAGCTCTACCCCGTGGCTGCTAAGACACAGCTTGGCCGGTGCTTCCAGGA  
 AACCATTGGGTACAGTGTGGCAAAGCGCTCCCTGGCCTTTCAAAGCAAGAGGACTGC  
 TGTGGAACTGTGGGTACCTCTGGGGCTTTAACAATGCCAGAAATGCCCAAGAAAC  
 CATCTTATCATGGATACAACCAATGATGGAATGCCTACCGGGTTATAAGCGGGTTAA  
 CAACACCTTTTGCCAAGATATTAATGAATGTGAGTACAAGGTGATGCCCTAATGGT  
 GAGTGTTTGAATACCATGGGCAGCTATCGATGTACCTGCAAAATAGGATTTGGGCCGG  
 ATCCTACCTTTTCAAGTTGTGTTCTGTATCCCTGTGATCTCGGAAGAGAAAGGGCC  
 CTGTTACCGACTTGTGAGTTCTGGAAAGACAGTGTATGTACCTCTGTCTGTTCACTC  
 ACCAAGCAGCTCTGCTGTTGAGTGTGGGCAAGGCTGGGCCACACTGTGAGAAATGTC  
 CCCTTCCAGGCACAGCTGCTTTTAAAGAAATCTGTCTGTGGGAATGGGTTATACGGT  
 TTCTGGCGTTTCATAGACGACAGGCCAATCCATCACCATGTAGGTAAAGGACCTGTATTT  
 GTCAAGCCAAAGAACACTCAACCTGTTGCTAAAAGTACTCATCTCCACCTCTCCACG  
 CCAAGGAAGAGCCAGTGGAGGCCCTGACCTTCTCCGGGAACACGGGGCCAGGAGTGC  
 GGAGCCAGAAGTGGCAACTGCACCCCTGAAAAGGAAATACCTTCATTGGATCAAGAG  
 AAAACCAAACTTGAGCCTGGTCAACCCAGCTGTCTCCAGGCATTTCCGCTATTTCATC  
 TGCATCCACAGTTTCCAGTAGTGATTGAAAAACATCACCTCCTGTGCTGTTGAAGT  
 AGCTCCTGAAGCTTCTACGTCTAGTGCCAGCCAAGTGATTGCTCCTACTCAAGTGACA  
 GAAATCAATGAATGTACTGTGAACCTGATATCTGTGGAGCAGGACACTGCATTAAAC  
 TACCAGTGAGATATACCTGTATATGCTACGAGGGCTACAGGTTTCAAGTCAACAGAG  
 GAAATGTGTGGATATTGATGAGTGTACTCAGGTCCAACACCTCTGCTCCAGGGCCGC  
 TGTGAAAACACCGAGGGAAGTTTCTGTGTCATTGCCCAGCAGGATTATGGCCAGTG  
 AGGAGGGTACTAATGCATAGATGTTGACGAATGCCTGAGGCCGACGCTCTGTGGGA  
 GGGGCACTGTGTCAATACTGTGGGGGCTTCCGGTGTGAATACTGTGACAGCGGGTAC  
 CGCATGACTCAGAGAGGCCGTTGTGAGGATATTGATGAATGTTGAATCCAAGCACTT  
 GTCCAGATGAGCAGTGTGTGAATTCTCTGGATCTTACCAGTGCGTTCCCTGCACAGA  
 AGGATTCGAGGCTGGAATGGACAGTGCCTTGATGTGGACGAGTGCCTGGAACCAAC  
 GTCTGCGCAAATGGTGATTGTTCCAACCTTGAAGGCTCCTACATGTGTTTCATGCCACA  
 AAGGCTATACCCGACTCCGGACCACAAGCACTGTAGAGATATTGATGAATGTGAGCA  
 AGGGAATCTATGTGTAACGGGCAGTGCAAAAATACCGAGGGCTCCTTCAGGTGCACC  
 TGTGGACAGGGGGTTACCAGCTGTGCGCAGCTAAAGACCAGTGTGAAGACATTGATG  
 AATGCCAGCACCGTCATCTCTGTGCTCATGGGCAGTGCAGGAACACTGAGGGCTCTTT  
 TCAATGTGTGTGACCAGGGTTACAGAGCATCTGGGCTTGAGAGCACTGTGAAGAT  
 ATCAATGAATGCTTGGAGGACAAGAGTGTGTCAGAGAGGAGACTGCATTAATACTG  
 CAGGGTCTATGATTGTACTTGTCCGGATGGATTTCAGCTAGATGACAATAAACATG  
 TCAAGATATTAATGAATGTGAACATCCAGGGCTCTGTGGTCCACAAGGGGAGTGCCTA  
 AACACAGAGGGTTCTTTCCATTGTGTCTGCCAGCAGGGTTTCTCAATCTCTGCAGATG  
 GCCGTACGTGTGAAGATGTGAATGAATGTGAATGCTCAGTGGGGTGTGTGGTGAAGC  
 CTTCTGTGAAAACGTGGAAGGGTCTTCTGTGCGTGTGTGCTGATGAAAACCAAGAG  
 TACAGCCCCATGACTGGGCAGTGCCTGCTCCCGACCTCCACAGATTTAGATGTAGATG  
 TAGATCAACCCAAAGAAGAAAAGAAAGAAATGCTACTATAATCTCAATGACGCCAGTCT  
 CTGTGATAATGTGTTGGCCCCAATGTACGAAACAAGAAATGCTGCTGTACATCAGGC  
 GCGGGATGGGGAGATAACTGCGAAATCTTCCCTGCCCGTCTTGGGAACTGCTGAGT  
 TCACTGAAATGTGTCCCAAGGGGAAAGGTTTTGTGCTGCTGGAGAAATCATCTTCTGA  
 AGCTGGTGGTGAGAAGTATAAAGATGCAGATGAATGCCTACTTTTGGACAAGAAATC  
 TGCAAAATAGTTTCTGTTTGAACACTCGGCCTGGGTATGAATGCTACTGTAAAGCAAG  
 GGACGTACTATGATCCTGTGAAACTGCAGTGCTTGTATGATGATGATGTCAAGACCC  
 CAGTAGTTGTATTGATGGCCAGTGTGTTAATACAGAGGGCTCTTACAACCTGCTTCTGT  
 ACTACCCCATGGTCTGTGATGCGTCAGAAAAAGATGTATACGACCGGCTGAGTCAA  
 ACGAACAAATAGAAGAACTGATGTCTACCAAGATTGTGCTGGGAACATCTGAGTGA  
 TGAATACGTGTGTAGCCGGCTCTTGTGGGCAAGCAGACAACGTACACTGAGTGCTGC  
 TGTCTGTATGGAGAGCCCTGGGGCATGCAGTGTGCCCTCTGCCCCCTGAAGGATTACG

	ATGACTATGCTCAGCTGTGTAACATCCCCGTGACGGGACGCCGGCAGCCATATGGACG GGACGCCTTGGTTGACTTCAGTGAACAGTATACTCCAGAAGCCGATCCCTACTTCATC CAAGACCGTTTTCTAAATAGCTTTGAGGAGTTACAGGCTGAGGAATGCGGCATCCTCA ATGGATGTGAAAATGGTCGCTGTGTGAGGGTCCAGGAAGGTTACACCTGCGATTGCTT TGATGGGTATCACTTGGATACGGCCAAGATGACCTGTGTGCGATGTAATGAATGCGAT GAGTTGAACAACCGGATGTCTCTCTGCAAGAATGCCAAGTGCATTAAACCCGATGGTT CCTACAAGTGTGTGTCTGCCAGGCTACGTGCCTTCTGACAAGCCAACTACTGCAC TCCGTTGAATACCGCTTGAATTTAGAGAAAGACAGTGACCTGGAGTGAAACAGAATC TACATAACCTAAGCCCATATACTCTGCACTGTGTAAAGGAAAAGGGAGAAATGTATTA TACTTGAGACATTGCACCTACCCCGAAGGCTGGAATACGGAAACAGCATGGAGTTG CAAGTCTCTGAAGACAATGAGAGGATTTAGGATGAGCCCGATAGGTGTGGCAGACCA AATGGACATTTCTCTAAAAAACAGTATATATAGTCTGTTCATATGTAATAATTCATG GAAGAGAGGTGGAACAGTCTGTTATTTTAAACAGAAGGTTGTATTATATGTTGTTT TGTGTTTTTACTATTGCTTGATTAAATTTGGCATTAAATAGTGGTGGAATATTTTA TATAATTTTCATTTTTTGGTTGTGCAGTTCCTTGGCTACTGTTTTCTTTTACTTCAG TTTTTTAAAAATCTCAAATGAAAAAGTCTTCGATACAATATTGTTAAGCTGTATTATA AGTATTGTTACACAGGGTTATGCAATCCCGGCTGGAGCATTTTGAATTCAAATT GTCTGTCTGTGGAGCAGGCAGTGATTTTGTTCAAAACCTTGATATACADTTTGGAG AAAAGTACTTTATATTTTCACTGTTTGTCTGATTTTAAATGTCCTGTTCTTAGCCAAGC TGCTAGCAGGTGTTAATTGGATCCCTTTCCTTCACTGAAATGGAAGAGTTTATAAGCT TACGTTAGTATTGTAATATGTAAAGTAAGCCCAACAAAAATTTTTAAAAATTTGATGA TCCCCAATATATCTACCATTGTATGTTAAATAAATCACCATTTTTGTAGAAAAAATTC TACCTGAGAGTAATTGTCAATGAGTACATGTGTATAAGTTGTATCCCACTCTCCCCAC TTTTATCTTTTCCAGTGGTCTTCTGTTAATGTAGTGTCTTTTACAAGTTAATCATTA ATTTGTTAGATCTTGTATGGGCTAAAAAATAAAAAAAAAA		
	ORF Start: ATG at 56	ORF Stop: TGA at 5093	
	SEQ ID NO: 170	1679 aa	MW at 182193.4kD
NOV37c, CG59819-03 Protein Sequence	MAGAWLRWGLLLWAGLLASSAHGRLRRIYVVHPGPGLAAGALPLSGPPRSRTFNVAL NARYSRSSAAAGAPSRASPGVPSERTRRTSKPGGAALQGLRPPPPPPPEPARPAVPGG QLHPNPGGHPAAAPFTKQGRQVVRSKVPQETQSGGSRQLVHQKQLQGVNVCGGRCC HGWSKAPGSQRCTKPSCVPPCQNGMCLRPQLCVCKPGTKGACETIAAQDTSPPVFG GQSPGAASSWGPPEQAAKHTSSKKADTLPRVSPVAQMTLLTKPKPSVGLPQQIHSQVT PLSSQSVVIHHGQTQEVYLVKPKYFPAQKGISGEQSTEGSFPLRYVDQVAAPFQLSNH TGRIKVVFTPSICKVTCTKGSCQNSCEKGNNTTLLISENGHAADTLTATNFRVVIHCLP CMNGGQCSSRDCKQCPNFTGKLCQIPVHGASVPKLYQHSQQPGKALGTHVIHSTHTL PLTVTSQQGVKVFPPNIIVNHVHPPEASVQIHQVSRIDGPTGQKTKEAQPGQSQVS YQGLPVQKTQTIHSTYSHQQVIPHVYPVAAKTQLGRCFQETIGSQCKALPGLSKQED CCGTVGTSWGFNKCQKCPKPSYHGYNQMMELPGYKRVNNTFCQDINECQLQGVCPN GECLNTMGSYRCTCKIGFGPDPTFSSCVPDPPIVSEEKGPYRLVSSGRQCMYPLSVH LTKQLCCCSVGKAGPHCEKCLPGTAAAFKEICPGMGYTVSGVHRRRPIHHHVKGKGPV FVKPKNTQPVAKSTHPPPLPAKEEPVEALTFSREHGARSAPPEVATAPPEKEIPSLDQ ETKLEPGQPQLSPGISAIHLHPQFPVVEIKTSPPVPVEVAPEASTSSASQVIAPTQV TEINECTVNPDICGAGHCINLPVRYTCICYEGYRFSEQQRKCVDIDECTQVQHLCSQG RCENTEGSFLCICPAGFMASEEGTNCIDVDECLRPDVCGEHCVNTVGAFRCEYCDSG YRMTQGRCEDIDECLNPSTCPDEQCVNSPGSYQCVPCTEGFRGWNQCLDVECLEP NVCANGDCSNLEGSYMCCHKGYTRTPDHKHCRIDECQGNLCVNGQCKNTEGSFRC TCGQGGYQLSAAKDQCEDIDECQHRHLCAHGQCRNTEGSFCVCDQGYRASGLGDHCE DINECLEDKSVCQRGDCINTAGSYDCTCPDGFQLDNKTCDINECEHPGLCGPQGE LNTGFSFHCVCQGFSSISADGRTCEDVNECELLSGVCGEAFCEVNEGSFLCVCADENQ EYSPMTGQCRSRTSTDLDVDVDQPKKEKKECYYNLNDASLCDNVLAPNVTKQECCTTS GAGWGDNCEIFPCPVLGTAEFTEMCPKKGFPVAGESSEAGGENYKDADECLLFQGE ICKNGFCLNTRPGYECYCKQGTYYDPVKLQCFDMDECQDPSSCIDGQCVNTEGYNCF CTHPMVLDASEKRCIRPAESNEQIEETDVYQDLCEWHLSDIEYVCSRLPLVGKQTYTEC CCLYGEAWGMQCALCPLKSDDYAQLCNIPTVGRROPYGRDALVDFSEQYTPPADPYF IQDRFLNSFEELQABECGLNGCENGRCVRVQEGYTCDCFDGYHLDTAKMTCVDVNEC DELNNRMSLCKNAKCINTDGSYKCLCLPGYVPSDKPNYCTPLNTALNLEKSDLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 37B.

<b>Table 37B. Comparison of NOV37a against NOV37b through NOV37c.</b>		
<b>Protein Sequence</b>	<b>NOV37a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>
NOV37b	19..1721	1561/1703 (91%)
	19..1668	1562/1703 (91%)
NOV37c	19..1721	1565/1704 (91%)
	19..1679	1565/1704 (91%)

Further analysis of the NOV37a protein yielded the following properties shown in Table 37C.

<b>Table 37C. Protein Sequence Properties NOV37a</b>	
<b>PSort analysis:</b>	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
<b>SignalP analysis:</b>	Likely cleavage site between residues 24 and 25

A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several  
5 homologous proteins shown in Table 37D.

<b>Table 37D. Geneseq Results for NOV37a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV37a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAR22461	Masking protein high polymer unit precursor MPU-P - Rattus rattus, 1712 aa. [JP04066597-A, 02-MAR-1992]	1..1721 1..1712	1525/1721 (88%) 1603/1721 (92%)	0.0
AAR14584	TGF beta 1 binding protein encoded by clone BPA 13 - Homo sapiens, 1355 aa. [WO9113152-A, 05-SEP-1991]	342..1721 16..1355	1324/1380 (95%) 1326/1380 (95%)	0.0
AAR53089	Human masking protein subunit hMPU-P - Homo sapiens, 845 aa. [JP06092995-A, 05-APR-1994]	742..1586 1..845	841/845 (99%) 841/845 (99%)	0.0
AAR53086	Human masking protein subunit hMPU-1 - Homo sapiens, 756 aa. [JP06092995-A, 05-APR-1994]	832..1586 2..756	752/755 (99%) 752/755 (99%)	0.0
AAR53087				0.0

	hMPU-2 - Homo sapiens, 752 aa. [JP06092995-A, 05-APR-1994]	1..752	749/752 (99%)	
--	---	--------	---------------	--

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37E.

Table 37E. Public BLASTP Results for NOV37a				
Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q00918	Latent transforming growth factor beta binding protein 1 precursor (Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP- 1) (Transforming growth factor beta-1 masking protein, large subunit) - Rattus norvegicus (Rat), 1712 aa.	1..1721 1..1712	1536/1721 (89%) 1611/1721 (93%)	0.0
O88349	LATENT TGF BETA BINDING PROTEIN - Mus musculus (Mouse), 1713 aa.	1..1720 1..1712	1523/1721 (88%) 1603/1721 (92%)	0.0
P22064	Latent transforming growth factor beta binding protein 1 precursor (Transforming growth factor beta-1 binding protein 1) (TGF-beta1-BP- 1) - Homo sapiens (Human), 1394 aa.	342..1721 16..1394	1369/1380 (99%) 1370/1380 (99%)	0.0
O35806	LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN - Rattus norvegicus (Rat), 1764 aa.	72..1705 75..1760	710/1748 (40%) 937/1748 (52%)	0.0
Q14767	LATENT TRANSFORMING GROWTH FACTOR-BETA-BINDING PROTEIN-2 (LTBP-2) - Homo sapiens (Human), 1821 aa.	74..1706 87..1818	693/1810 (38%) 919/1810 (50%)	0.0

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37F.

Table 37F. Domain Analysis of NOV37a			
Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF: domain 1 of 18	191..218	15/47 (32%) 19/47 (40%)	0.0056



EGF: domain 2 of 18	403..430	15/47 (32%) 23/47 (49%)	0.00014
wap: domain 1 of 1	385..433	12/57 (21%) 30/57 (53%)	9.3
TB: domain 1 of 4	566..609	15/48 (31%) 41/48 (85%)	6e-13
EGF: domain 3 of 18	630..665	14/47 (30%) 27/47 (57%)	1e-05
Keratin_B2: domain 1 of 1	578..717	40/180 (22%) 64/180 (36%)	1.5
TB: domain 2 of 4	687..728	25/47 (53%) 40/47 (85%)	1.1e-21
Arthro_defensin: domain 1 of 1	874..901	9/37 (24%) 18/37 (49%)	8.4
EGF: domain 4 of 18	877..913	15/47 (32%) 27/47 (57%)	1.8e-05
EGF: domain 5 of 18	919..955	15/47 (32%) 27/47 (57%)	6e-05
granulin: domain 1 of 2	942..957	6/16 (38%) 12/16 (75%)	0.57
EGF: domain 6 of 18	961..996	15/47 (32%) 22/47 (47%)	7.9
EGF: domain 7 of 18	1002..1036	13/47 (28%) 27/47 (57%)	50
EGF: domain 8 of 18	1042..1077	15/47 (32%) 24/47 (51%)	0.00066
EGF: domain 9 of 18	1083..1118	16/47 (34%) 30/47 (64%)	0.00019
EGF: domain 10 of 18	1124..1159	14/47 (30%) 28/47 (60%)	0.00026
EGF: domain 11 of 18	1165..1200	14/47 (30%) 26/47 (55%)	0.0071
EGF: domain 12 of 18	1206..1242	13/47 (28%) 27/47 (57%)	0.00073
granulin: domain 2 of 2	1226..1244	10/19 (53%) 15/19 (79%)	20
EGF: domain 13 of 18	1248..1284	13/47 (28%) 25/47 (53%)	0.00063
EGF: domain 14 of 18	1290..1327		0.0037



		27/47 (57%)	
TB: domain 3 of 4	1357..1400	24/47 (51%) 36/47 (77%)	2e-18
EGF: domain 15 of 18	1428..1465	14/47 (30%) 27/47 (57%)	0.014
EGF: domain 16 of 18	1471..1506	15/47 (32%) 29/47 (62%)	1.2e-05
TB: domain 4 of 4	1534..1576	18/47 (38%) 40/47 (85%)	8.6e-18
EGF: domain 17 of 18	1625..1660	16/47 (34%) 26/47 (55%)	0.0004
EGF: domain 18 of 18	1666..1705	16/49 (33%) 31/49 (63%)	5.8e-06

**EXAMPLE 38.**

The NOV38 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis			
	SEQ ID NO: 171		1034 bp
NOV38a, CG59685-01 DNA Sequence	GCGGCCGCCCGCGGGCTCCTGGAACCCCGGTTGCGGGCGATGCCAGCCACCCAGCG AAGCCGCCGAGTTTCACTGCTTGGATAATTTGAAAGTACAATAGTTGGTTTCCCTGTC CACCCGCCCACTTTCGCTTGCATCACAGCACGCCTATCGGATGTGAGAGGAGAAGTC CCGCTGCTCGGGCACTGTCTATATACGCCTAACACCTACATATATTTTAAAAACATTA AATATAATTAAACAATCAAAAGAAAGAGGAGAAAGGAAGGGAAGCATTTACTGGGTTACT ATGCACTTGCGACTGATTTCTTGGCTTTTTATCATTTTGAACTTTATGGAATACATCG GCAGCCAAAACGCCTCCCGGGGAAGGCGCCAGCGAAGAATGCATCCTAACGTTAGTCA AGGCTGCCAAGGAGGCTGTGCAACATGCTCAGATTACAATGGATGTTTGTCAATGAAG CCCAGACTATTTTTGTCTGGAAGAATTGGCATGAAGCAGATTGGAGTATGTCTCT CTTATGTCCTAAGTGGATATTATGGAACCTCGATATCCAGATATAAATAAGTGTACAAG TAAGTGCCCAACACGAAAAGCTGACTGTGATACCTGTTTCAACAAAAATTTCTGCACA AAATGTAAAAGTGGATTTTACTTACACCTTGGAAAGTGCCCTTGACAATTGCCGAGAAG GGTTGGAAGCCAACAACCATATGAGAGTGTGTGAGTTCAGTGCAGTGTGAGGTGAGT TGAATGGAATCCTTGGAGTCCATGCACGAAGAAGGGAAAAACATGTGGCTTCAAAAGA GGGACTGAAACACGGGTCGAGAAATAATACAGCATCCTTCAGCAAGGGTAACCTGT GTCCCCCAACAATGAGACAAGAAAGTGTACAGTGCAAAGGAAGAAGTGTGAGAAGGG AGAACGAGGTACAATCATAATAACAAAATGTGCTTGTGTTGAATCCTCATAATCTGTG CATTTTTCATTTTATTTCTTATGAAACACTTGGCATTATCTTTTCATGC		
	ORF Start: ATG at 291		ORF Stop: TGA at 1008
	SEQ ID NO: 172		239 aa      MW at 27062.1kD
NOV38a, CG59685-01 Protein Sequence	MHLRLISWLFIIILNFMEYIGSQNASRGRRRQRMHPNVSQGCQGGCATCSYNGCLSK PRLFFALERIGMKQIGVCLSSCPSGYGTRYPDINKCTSKCPHEKADCDTCFNKNFCT KCKSGFYHLGLKCLDNCEPGLLEANNHTMECVSVHCEVSEWNPWSPCTKKGKTCGFKR GTETRVREIIQHPSAKGNLCPPTNETRKCTVQRKKCKQGERGTIIITKACALNPHNLL HFSFYFL		
	SEQ ID NO: 173		585 bp
NOV38b,	GGATCCCAAAACGCCTCCCGGGGAAGGCGCCAGCGAAGAATGCATCCTAACGTTAGTC AAGGCTGCCGAGGAGGCTGTGCAACATGCTCAGATTACAATGGATGTTTGTCAATGTA		

175070296 DNA Sequence	GCCAGACTATTTTCTGCTCTGGAAGAATTGGCATGAAGCAGATTGGAGTATGTCTCTCTTCATGTCCAAGTGGATATTATGGAAGCTGATATCCAGATATAAATAAGTGACAAATGCAAAGCTGACTGTGATACCTGTTTCAACAAAAATTTCTGCACAAAATGTAAAAGTGGATTTTACTTACACCTTGGAAAGTGCCTTGACAATTGCCAGAAGGGTTGGAAGCCAACAACCATACTATGGAGTGTGTGAGTATTGTGCACTGTGAGGTGAGTGAATGGAATCCTTGAGTCCATGCACGAAGAAGGAAAAACATGTGGCTTCAAAGAGGGACTGAAACACGGGTCCGAGAAATAATACAGCATCCTTCAGCAAAGGGTAACCTGTGTCCCCCAACAATGAGACAAGAAAGTGTACAGTGCAAAGGAAGAAGTGTGAGAAGGGAGAACGAGGTCTCGAG		
	ORF Start: GGA at 1	ORF Stop: at 586	
	SEQ ID NO: 174	195 aa	MW at 21781.8kD
NOV38b, 175070296 Protein Sequence	GSQNASRGRRRQRRMHPNVSQGRGGCATCSDYNGCLSKPRLFFALERIGMKQIGVCLSSCPSGYGTRYDPINKCTKCKADCDTCFNKNFCTKCKSGFYHLHLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPWSPCTKKGKTCGFKRGTETRVREIIQHPSAKGNLCPPTNETRKCTVQRKKCQKGERGLE		
	SEQ ID NO: 175	585 bp	
NOV38c, 175070324 DNA Sequence	GGATCCCAAAACGCCTCCCGGGGAAGGCGCCAGCGAAGAATGCATCCTAACGTTAGTCAAGGCTGCCAAGGAGGCTGTGCAACATGCTCAGATTACAATGGATGTTTGTCTATGTAAAGCCCAGACTATTTTCTGCTCTGGAAGAATTGGCATGAAGCAGATTGGAGTATGTCTCTCTTCATGTCCAAGTGGATATTATGGAAGCTGATATCCAGATATAAATAAGTGACAAATGCAAAGCTGACTGTGATACCTGTTTCAACAAAAATTTCTGCACAAAATGTAAAAGTGGATTTTACTTACACCTTGGAAAGTGCCTTGACAATTGCCAGAAGGGTTGGAAGCCAACAACCATACTATGGAGTGTGTGAGTATTGTGCACTGTGAGGTGAGTGAATGGAATCCTTGAGTCCATGCACGAAGAAGGAAAAACATGTGGCTTCAAAGAGGGACTGAAACACGGGTCCGAGAAATAATACAGCATCCTTCAGCAAAGGGTAACCTATGTCCCCCAACAATGAGACAAGAAAGTGTACAGTGCAAAGGAAGAAGTGTGAGAAGGGAGAACGAGGTCTCGAG		
	ORF Start: GGA at 1	ORF Stop: at 586	
	SEQ ID NO: 176	195 aa	MW at 21753.8kD
NOV38c, 175070324 Protein Sequence	GSQNASRGRRRQRRMHPNVSQCGGCGCATCSDYNGCLSKPRLFFALERIGMKQIGVCLSSCPSGYGTRYDPINKCTKCKADCDTCFNKNFCTKCKSGFYHLHLGKCLDNCPEGLEANNHTMECVSIHVCEVSEWNPWSPCTKKGKTCGFKRGTETRVREIIQHPSAKGNLCPPTNETRKCTVQRKKCQKGERGLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 38B.

Table 38B. Comparison of NOV38a against NOV38b through NOV38c.		
Protein Sequence	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV38b	20..216	179/197 (90%)
	1..193	180/197 (90%)
NOV38c	20..216	180/197 (91%)
	1..193	180/197 (91%)

Further analysis of the NOV38a protein yielded the following properties shown in Table 38C.

Table 38C. Protein Sequence Properties NOV38a
---

PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 22 and 23

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 38D.

Table 38D. Geneseq Results for NOV38a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE13170	Human SCR-1 related protein - Unidentified, 292 aa. [WO200177169-A2, 18-OCT-2001]	1..216 1..212	211/216 (97%) 211/216 (97%)	e-130
AAE13168	Human stem cell growth factor-like protein #4 - Homo sapiens, 272 aa. [WO200177169-A2, 18-OCT-2001]	1..216 1..212	211/216 (97%) 211/216 (97%)	e-130
AAE13163	Human secreted protein from clone DA228_6 - Homo sapiens, 265 aa. [WO200177169-A2, 18-OCT-2001]	1..216 1..212	211/216 (97%) 211/216 (97%)	e-130
AAE13150	Human stem cell growth factor-like protein #2 - Homo sapiens, 272 aa. [WO200177169-A2, 18-OCT-2001]	1..216 1..212	211/216 (97%) 211/216 (97%)	e-130
AAM78328	Human protein SEQ ID NO 990 - Homo sapiens, 272 aa. [WO200157190-A2, 09-AUG-2001]	1..216 1..212	211/216 (97%) 211/216 (97%)	e-130

In a BLAST search of public sequence databases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38E.

Table 38E. Public BLASTP Results for NOV38a				
Protein Accession Number	Protein/Organism/Length	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BXY4	THROMBOSPONDIN - Homo sapiens (Human), 272 aa.	1..216 1..212	211/216 (97%) 211/216 (97%)	e-129
CAD10541	SEQUENCE 12 FROM PATENT	2..216 3..213	210/215 (97%) 210/215 (97%)	e-129

	273 aa.			
Q96K87	CDNA FLJ14440 FIS, CLONE HEMBB1000915, WEAKLY SIMILAR TO SUBTILISIN-LIKE PROTEASE PACE4 PRECURSOR (EC 3.4.21.-) - Homo sapiens (Human), 292 aa.	1..216 1..212	209/216 (96%) 209/216 (96%)	e-127
Q9CSB2	2810459H04RIK PROTEIN - Mus musculus (Mouse), 217 aa (fragment).	1..216 1..212	196/216 (90%) 201/216 (92%)	e-120
CAD10542	SEQUENCE 31 FROM PATENT WO0177169 - Mus musculus (Mouse), 279 aa.	1..216 1..214	197/218 (90%) 202/218 (92%)	e-119

Pfam analysis predicts that the NOV38a protein contains the domains shown in the Table 38F.

Table 38F. Domain Analysis of NOV38a			
Pfam Domain	NOV38a Match Region	Identities/ Similarities for the Matched Region	Expect Value
GASA: domain 1 of 1	1..100	19/121 (16%) 59/121 (49%)	6
EB: domain 1 of 1	80..129	14/64 (22%) 32/64 (50%)	6.3
Plexin_repeat: domain 1 of 1	98..147	10/72 (14%) 31/72 (43%)	3.2
tsp_1: domain 1 of 1	155..210	20/61 (33%) 46/61 (75%)	0.002

#### EXAMPLE 39.

The NOV39 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 39A.

Table 39A. NOV39 Sequence Analysis		
	SEQ ID NO: 177	1020 bp
NOV39a, CG57167-01 DNA Sequence	ATGGGGTCCCCAGAGTCATCTGCTCTGCCTCTTTGGGGCTGCGCTCTGCCTGACAG GGTCCCAAGCCCTGCAGTGCTACAGCTTTGAGCACACCTACTTTGGCCCTTTGACCT CAGGGCCATGAAGCTGCCAGCATCTCCTGTCCTCATGAGTGCTTTGAGGCTATCCTG TCTCTGGACACCGGTATCGCGCGCCGGTGACCCTGGTGCGGAAGGGCTGCTGGACCG GGCCTCCTGCGGGCCAGACGCAATCGAACC CGGACGCGCTGCCGCCAGACTACTCGGT GGTGCGGGCTGCACAACTGACAAATGCAACGCCACCTCATGACTCATGACGCCCTC CCCAACCTGAGCCAAGCACCCGACCCGCCGACGCTCAGCGGCGCCGAGTGCTACGCCT GTATCGGGGTCCACCAGGATGACTGCGCTATCGGCAGGTCCCGACGAGTCCAGTGTC CCAGGACCAGACCGCTGCTTCCAGGCAATGGCAGAATGACAGTTGCAATTTCTCA	

	GTCCCTGTGTACATCAGAACCTGCCACCGGCCCTCCTGCACCACCGAGGGCACCACCA GCCCCTGGACAGCCATCGACCTCCAGGGCTCCTGCTGTGAGGGGTACCTCTGCAACAG GAAATCCATGACCCAGCCCTTCACCAGTGCTTCAGCCACCACCCCTCCCCGAGCACTA CAGGTCCTGGCCCTGCTCCTCCCAGTCCTCCTGCTGAAAAACACACAAGGCAAAGTTC AGCGAGGTGAAATTCTCCAAGCTATAAAGATCAGGGAAAGACTTCTCTGGAGGAATTCAC CCTTGAGCAAAATCCTAAAGGATCAATAGTAGCTGGCAAAAAGAAGCAGGAGGAAGCG CATTCTAGGCCATGTGACAAGGGCTTCAGGTGTCTTTACATCCTGACATACAAGGGGA AGCTGGATGTCTTCATTATCCTTCACATTACTGAGCACCTACTATGTGCAAGGCAC TGTTCAGTTGCTGGGCATGCAGCAGGGAACATA		
	ORF Start: ATG at 1	ORF Stop: TAA at 1018	
	SEQ ID NO: 178	339 aa	MW at 36956.0kD
NOV39a, CG57167-01 Protein Sequence	MGVPFVILLCLFGAALCLTGSQALQCYSFEHTYFGPFDLRAMKLPSISCPHECFEAIL SLDTGYRAPVTLVRKGCWTGPPAGQTQSNPDALPPDYSVVRGCTTDKCNHMLTHDAL PNLSQAPDPPTLSGAECYACIGVHQDDCAIGRSRRVQCHQDQTACFQNGRMTVGNFS VPVYIRTCHRPSCCTTEGTTSPWTAIDLQGSCEGYLCNRKSMTQPTTSASATTPRAL QVLALLLPVLLKNTQGVQGEILQAIKIREDFLEETLEQNPKGSIVAGKKKQEEA HSRPCDKGFRCLYILTYKGKLDVFIHPSHLLSTYYVQGTVPVAGHAAGN		

Further analysis of the NOV39a protein yielded the following properties shown in Table 39B.

Table 39B. Protein Sequence Properties NOV39a	
PSort analysis:	0.8200 probability located in outside; 0.4575 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 24 and 25

- A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 39C.

Table 39C. Geneseq Results for NOV39a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU29261	Human PRO polypeptide sequence #238 - Homo sapiens, 251 aa. [WO200168848-A2, 20-SEP-2001]	1..244 3..246	243/244 (99%) 244/244 (99%)	e-148
AAB31206	Amino acid sequence of human polypeptide PRO4356 - Homo sapiens, 251 aa. [WO200077037-A2, 21-DEC-2000]	1..244 3..246	243/244 (99%) 244/244 (99%)	e-148
AAB18919	A novel polypeptide designated PRO4356 - Homo sapiens, 251 aa. [WO200056889-A2, 28-SEP-2000]	1..244 3..246	243/244 (99%) 244/244 (99%)	e-148

ABB16784	Human nervous system related polypeptide SEQ ID NO 5441 - Homo sapiens, 252 aa. [WO200159063-A2, 16-AUG-2001]	1..244 4..247	240/244 (98%) 240/244 (98%)	e-146
AAM24186	Human EST encoded protein SEQ ID NO: 1711 - Homo sapiens, 253 aa. [WO200154477-A2, 02-AUG-2001]	1..244 3..248	233/246 (94%) 234/246 (94%)	e-137

In a BLAST search of public sequence databases, the NOV39a protein was found to have homology to the proteins shown in the BLASTP data in Table 39D.

Table 39D. Public BLASTP Results for NOV39a				
Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96DR2	CDNA FLJ30469 FIS, CLONE BRAWH1000037, WEAKLY SIMILAR TO UROKINASE PLASMINOGEN ACTIVATOR SURFACE RECEPTOR PRECURSOR - Homo sapiens (Human), 208 aa.	42..244 1..203	202/203 (99%) 203/203 (99%)	e-122
Q9D7Z7	2210003I03RIK PROTEIN - Mus musculus (Mouse), 256 aa.	1..244 1..244	175/244 (71%) 201/244 (81%)	e-109
Q9UJ74	HYPOTHETICAL 36.0 KDA PROTEIN (C4.4A PROTEIN) - Homo sapiens (Human), 346 aa.	20..212 27..222	62/203 (30%) 96/203 (46%)	6e-15
O55162	METASTASIS-ASSOCIATED GPI-ANCHORED PROTEIN - Rattus norvegicus (Rat), 352 aa.	9..232 19..242	70/235 (29%) 103/235 (43%)	1e-14
O95274	GPI-ANCHORED METASTASIS-ASSOCIATED PROTEIN HOMOLOG - Homo sapiens (Human), 346 aa.	20..212 27..222	62/203 (30%) 95/203 (46%)	1e-14

PFam analysis predicts that the NOV39a protein contains the domains shown in the Table 39E.

Table 39E. Domain Analysis of NOV39a			
Pfam Domain	NOV39a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			



**EXAMPLE 40.**

The NOV40 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 40A.

Table 40A. NOV40 Sequence Analysis		
	SEQ ID NO: 179	6797 bp
NOV40a, CG59841-01 DNA Sequence	GAGGAGCCCCCAGCTCTGAGAGTGGGGCGCAGAGCCGAGCCCCGGGCCATGCCTCC GCTGCCGCTGGCGCGGACACCCGGCAGCCGCTGGCGCTCCCTGCTGGTGCAGGGC TTCATGGTGCCTGCAACGCTGCCTGATCTGCTGGCCACCGCCACGCTCGGCTTCG CGGTGCTGCTGTCTCAACAAGTGTAAACCCGGGACCACTTCACTCCAGTGCCTCC GACGCTCTGATCCATGCCTCGGGGTGCAGTGTGCATTTGGGGCGACGTGTGCTGTG AAGAACGGGCAGGCAGCGTGTGAATGCCTGCAGGCGTGTCTGAGCCTCTACGATCCTG TGTGCGGCAGCGACGGCGTCACATACGGCAGCGCGTGCAGCTGGAGGCCACGGCCTG TACCCTCGGGCGGGAGATCCAGGTGGCGCGCAAAGGACCTGTGGTTCGCGGACCCCC TGCTCCAACGTGACCTGCAGCTTCGGCAGCACCTGTGCGCGCTCGGCCGACGGCTGA CGGCTCGTGCTGTGCCCCGCGACCTGCCGTGGCGCCCCGAGGGGACCGTCTGCGG CAGCGACGGCGCGACTACCCCGGCGAGTGCCAGCTCCTGCGCCGCGCTGCGCCCGC CAGGAGAATGTCTTCAAGAAGTTCGACGGCCCTTGTGACCCCTGTGAGGGCGCCCTCC CTGACCCGAGCCGAGCTGCCGTGTGAACCCGCGCACGCGCGCCCTGAGATGCTCCT ACGGCCCGAGAGCTGCCCTGCCCGGAGGCGCCAGTGTGTGGGGACGACGGAGTCAAC TACGAAAACGACTGTGTATGGGCGGATCGGGGGCGCCCGGGGTCTCTCTCGCAGA AAGTGCCTCCGGCCAGTGCCAGGGTGCAGACAGTGCCCGGAGCCCTGCCGGTTCAA TGCCGTGTGCTGTCCCGCGTGGCGCTCCCGCTGCTCCTGCGACCGCTCACCTGT GACGGGGCTACAGGCCGTGTGTGCCAGGACGGGCGCACGTATGACAGTGATTGCT GGCGGCAGCAGGCTGAGTGCCGGCAGCAGCGTGCCATCCCGAGCAAGCACAGGGCCC GTGTGACCAAGCCCCGTCCCATGCTCGGGGTGCAGTGTGCATTTGGGGCGAGCTGT GCTGTGAAGAACGGGCAGGCAGCGTGTGAATGCCTGCAGGCGTGTCTGAGCCTCTACG ATCCTGTGTGCGGCAGCGACGGCGTCACATACGGCAGCGGTGCGAGCTGGAGGCCAC GGCCTGTACCTCGGGCGGGAGATCCAGGTGGCGGACCGCTGCGGGCAGTGCCGCTTT GGAGCCCTGTGCGAGGCCGAGACCGGGCGCTGCGTGTGCCCTCTGAATGCGTGGCTT TGGCCAGCCCCGTGTGTGCTCCGACGGGCACACGTACCCAGCGAGTGCATGTGCA CGTGACGCCTGCACACACAGATCAGCCTGCAGTGGCCTCAGCTGGACCCCTGTGAG ACCTGTGGAGATGCCGTGTGTGCTTTTGGGGCTGTGTGCTCCGACGGGCAGTGTGTG GTCCCCGCTGTGAGCACCCCGCCGCGCCCGTGTGTGGCAGCGACGGTGTACCTA CGGCAGTGCTGCGAGCTACGGGAAGCCGCTGCCTCCAGCAGACACAGATCGAGGAG GCCGGGCAGGGCCGTGCGAGCAGGCCGAGTGCCTTCCGGAGGCTCTGGCTCTGGGG AGGACGGTGACTGTGAGCAGGAGCTGTGCCGCGAGCGCGTGGCATCTGGGACGAGGA CTCGGAGGACGGGCCGTGTGTCTGTGACTTACGCTGCCAGAGTGTCCAGGCAGCCCCG GTGTGCGGCTCAGATGGGGTCACCTACAGCACCGAGTGTGAGCTGAAGAAGGCCAGGT GTGAGTCACAGCGAGGGCTTACGTAGCGGCCAGGGAGCCTGCCGAGGCCACCTT CGCCCCGCTGCCGCTGTGGCCCCCTTACACTGTGCCAGACGCCCTACGGCTGCTGC CAGGACAATATACCGCAGCCCGGGCGTGGGCTGGCTGGCTGCCCGAGTGCCTGCC AGTGCAACCCCATGGCTCTTACGGCGGCACCTGTGACCCAGCCACAGGCCAGTGCTC CTGCCGCCAGGTGTGGGGGCTCAGGTGTGACCGCTGTGAGCCTGGCTTCTGGAAC TTTCGAGGCATCGTACCGATGGCCGAGTGGCTGTACACCTCGCAGCTGTGATCCCC AAGGCGCCGTGCGGGATGACTGTGAGCAGATGACGGGGCTGTGCTCGTGAAGCCCGG GGTGGCTGGACCAAGTGTGGGCAGTGTCCAGACGGCCGTGCCCTGGGCCCGCGGGC TGTGAAGCTGACGCTTCTGCGCTGCGACCTGTGCGGAGATGCGCTGTGAGTTCGGTG CGCGGTGCGTGGAGGAGTCTGGCTCAGCCACTGTGTCTGCCCGATGCTCACCTGTCC AGAGGCCAACGCTACCAAGGTCTGTGGGTACAGTGGAGTACATACGGCAACGAGTGT CAGCTGAAGACCATCGCCTGCCGCCAGGGCCTGCAAATCTCTATCCAGAGCCTGGGCC CGTGCCAGGAGGCTGTGTCTCCAGCACTACCCGACATCTGCCTCCGTGACTGTGAC CACCCAGGGCTCCTCCTGAGCCAGGCACTGCCGGCCCCCCCCGGCGCCCTCCCCCTG GCTCCAGCAGTACCGCACACAGCCAGACACCCCTCCGCCCTCATCGCGACCTCGGA CCACTGCCAGCGTCCCGAGGACACCGTGTGGCCCGTGTGACGGTCCCCCGACGGC ACCTCCCCCTGCACCCAGCTGGTGGCGTCCGCTTTGGTGAATCTGGCAGCACTGAT GGAAGCAGCGATGAGGAAGTGCAGGGGACAGGAGGCCAGTGGGGGTGGCTCTGGGG	

GGCTCGAGCCCTTGGAGGGCAGCAGCGTGGCCACCCCTGGGCCACCTGTCTGAGAGGGC  
 TTCTCTGCTACAACCTCCGCGTTGGGCTGCTGCTCTGATGGGAAGACGCCCTCGCTGGAC  
 GCAGAGGGCTCCAAGTCCCGCCACCAAGGTGTTCCAGGGCGTCTGGAGCTGGAGG  
 GCGTCGAGGGCCAGGAGCTGTTCTACACGCCGAGATGGCTGACCCCAAGTCAGAACT  
 GTTCGGGGAGACAGCCAGGAGCATTGAGAGCACCTTGGACGACCTCTTCCGGAATTCA  
 GACGTCAAGAAGGATTTCCGGAGTGTCCGCTTGGGGACCTGGGGCCCGGCAATCCG  
 TCCGCGCCATTGTGGATGTGCACTTTGACCCACCAAGCCTTCAGGGCACCACGCT  
 GGCCCGGGCCCTGCTCCGCGAGATCCAGGTGTCCAGGCGCCGGTCTTGGGGGTGAGG  
 CGGCCGCTGCAGGAGCAGTGCATTATGGACTTTGACTGGTTTCTGCGTTTATCA  
 CGGGGGCCACGTAGGAGCCATTGCTGCGGGAGCCACGGCCAGAGCCACCACCTGCATC  
 GCGCTGCGCTCTCTGCTGTGACCCCTCGGGCCCCGACCCAGTCACACAAGCCAG  
 CCCGTGCGCAAGACCAGGACGCCCCCACCACAGTCCGCCCCCACCACCTGCCCCCA  
 GCCGTGTGCCCGAGCTCGGCCCCCGCCCCCAGCAGCCTTCCAAAGCCCTGTGACTC  
 ACAGCCCTGCTTCCACGGGGGACCTGCCAGGACTGGGCATTGGGCGGGGGCTTACC  
 TGCAGCTGCCCGCAGGCAGGGGAGGCGCCGTCTGTGAGAAGGTGCTTGGCGCCCTG  
 TGCCGCGCTTCAGGGGCGCTCCTTCTTGGCTTCCCCACCTCCGCGCCTACCACAC  
 GCTGCGCCTGGCACTGGAATTCCGGGCGCTGGAGCCTCAGGGGTGCTGTGTACAAT  
 GGCAACGCCCGGGCAAGGACTTCTGGCATTGGCGCTGTAGATGGCCGCTGCAGCAG  
 TCAGGTTTGACACAGGTTCCGGGGCCGGGCTGTGACCACTGCCGTGCCGTAGAGCC  
 GGGCAGTGGCACCCCTGGAGCTGTCCCGCACTGGCGCCGGGGCACCCCTCTCGGTG  
 GATGTGAGACCCCTGTTCTGGGCGAGAGTCCAGTGGCACCGACGCGCTCAACCTGG  
 ACACAGACCTCTTGTGGGCGGCTACCCGAGGACCAGGCTGCCGTGGCGCTGGAGCG  
 GACCTTCGTGGGCGCCGCTGAGGGGGTGATCCGTTTGTGTGCTCAACAGCCTGACAG  
 CGCTGGAGCTTGGCATTGGGCGGGGGCTGCCACCCAGGCTCTGGCGTGGGCGAGT  
 GCGGGGACACCCCTGCCCTGCCAACCCCTGCCATGGCGGGGCCCATGCCAGAACCT  
 GGAGGCTGGAAGGTTCCATTGCCAGTGCCCGCCCGGCGCGCTCGGAGCAACCTGTGCC  
 GATGAGAAGAGCCCTGCCAGCCCAACCCCTGCCATGGGGCGGCGCCTGCCGTGTGC  
 TGCCCGAGGGTGGTGTCTAGTGGAGTGCCCTTGGGGCGTGGGGCACCTTCTGCCA  
 GACAGCTCGGGGCGAGGACGCTCTGGGCGCTTCTGGCTGACTGCTCAACGCTTCTCC  
 CACCTGGAGCTGAGAGGCTGCACACCTTTGCACGGGACCTGGGGGAGAAGATGGCGC  
 TGGAGGTGCTGTTCTTGGCACGAGGCCCCAGCGGCTCTGCTCTACAACGGGCGAGAA  
 GACGGACGGCAAGGGGGACTTCGTGTGCTGGCACTGCGGGACCGCCGCTGGAGTTT  
 CGCTACGACCTGGGCAAGGGGGCAGCGGTATCAGGAGCAGGGAGCCAGTCACCTGG  
 GAGCTGGACAGGGTCTCACTGGAGCGAAACGGCCGCAAGGGTGCCCTGCGTGGG  
 CGACGCCCCCGTGTGTTGGGGAGTCCCGGTTCCGCACACCGTCTCAACCTGAAG  
 GAGCCGCTCTACGTAGGGGGCGCTCCCGACTTCAGCAAGCTGGCCCGTGTGCTGCCG  
 TGTCTCTGGCTTCGACGGTGCCATCCAGCTGGTCTCCCTCGGAGGCGCCAGCTGCT  
 GACCCCGGAGCAGTGTGCGGCAGGTGGACGTACGTCCTTTGCAGGTACCCCTGC  
 ACCCGGGCTCAGGCCACCCCTGCCCTCAATGGGGCTCCTGCGTCCCGAGGGAGGCTG  
 CCTATGTGTGCTGTGTTCCGGGGGATTCTCAGGACCGCACTGCGAGAAGGGGCTGGT  
 GGAGAAGTCAGCGGGGACGTGGATACCTTGGCCTTTGACGGGCGGACCTTTGTGAG  
 TACCTCAACGCTGTGACCGAGAGCGAGAAGGCACTGCAGAGCAACCACTTTGAACCTGA  
 GCCTGCGCACTGAGGCCACGACAGGGGCTGGTGTCTGGAGTGGCAAGGCCACGGAGCG  
 GGCAGACTATGTGGCACTGGCCATTGTGGACGGGCACCTGCAACTGAGCTACAACCTG  
 GGCTCCAGCCCGTGGTGTGCTGCGTTCCACCGTCCCGTCAACACCAACCGCTGGTTGC  
 GGGTCTGGCACATAGGGAGCAGAGGGAAGGTTCCCTGCAGGTGGGCAATGAGGCCCC  
 TGTGACCGGCTCTCCCGCTGGGCGCCACGAGCTGGACACTGATGGAGCCCTGTGG  
 CTTGGGGGCTGCCGAGCTGCCCGTGGGCCCAGCACTGCCCAAGGCTACGGCACAG  
 GCTTTGTGGGCTGCTTGGGGATGTGGTGGTGGGCGGCAACCGCTGCACCTGCTGGA  
 GGACGCGCTACCAAGCCAGAGCTGCGGCCCTGCCCCACCCCATGAGCTGGCACACAGA  
 GCCCCGCGCCGCTGTAATTATTTTCTATTTTGTAACTTGTGTGTTTTGTATATGA  
 TTTTCTGCTGAGTGTGGCCGGAGGGACTGTGGCCCGGCTCCTTCCGTCCAGG  
 CAGCCGTGCTGCAGACAGACCTAGTGTGAGGGATGGACAGGCGAGGTGGCAGCGTGG  
 AGGGCTCGGCGTGGATGGCAGCCTCAGGACACACACCCCTGCCTCAAGGTGCTGAGCC  
 CCCGCTTGCAGTGGCCTGCCCCACGGTGTCCCGCGGGGAAGCAGCCCCGCTCTCT  
 GAATCACCCCTCGTCCGTGAGGCGGACTCGTGTCCCAAAAGGAAGGGGCTGTGAG  
 GTCTGATGGGGCCCTTCTCCGGGTGACCCACAGGGCCTTTCCAAGCCCTATTGTA  
 GCTGCTCCTTCTGTGTGTGCTCTGGACCTGCCTCGGCTCCTGCGCAATACTGTG  
 ACTTCCAAACAATGTTACTGTGGGCACAGCTGTGCTTGTCCCGTGTGCTGCTGCGC  
 CAGCCCCAGGCTGCTGAGGAGCAGAGGCCAGACCAGGGCCGATCTGGGTGTCTGACC  
 CTCAGCTGGCCCTGCCGACCCCTGGACATGACCGTATCCCTCTGCCACACCCAG

	GCCCTGCGAGGGGCTATCGAGAGGAGCTCACTGTGGGATGGGGTTGACCTCTGCCGCC TGCCTGGGTATCTGGGCCTGGCCATGGCTGTGTCTTCATGTGTTGATTTTATTGAC CCCTGGAGTGGTGGGTCTCATCTTTCCCATCTCGCCTGAGAGCGGCTGAGGGCTGCCT CACTGCAATCCTCCCCACAGCGTCAGTGAAAGTCGTCTTGTCTCAGAATGACCAGG GGCCAGCCAGTGTCTGACCAAGGTCAAGGGGCAGGTGCAGAGGTGGCAGGGATGGCTC CGAAGCCAGAA		
	ORF Start: ATG at 51	ORF Stop: TGA at 5844	
	SEQ ID NO: 180	1931 aa	MW at 201789.3kD
NOV40a, CG59841-01 Protein Sequence	MPPLPLARDTRQPPGASLLVRGFMVPCNACLILLATATLGFVALLFLNNCKPGTHFTP VPPTPPDPCLGVQCAFGATCAVKNGQAACECLQACSSLYDPVCGSDGVTYGSACELEA TACTLGREIQVARKGPCSRDPSCSNVTCFSGSTCARSADGLTASCLCPATCRGAPEGT VCGSDGADYPGECQLLRACARQENVFKKFDGPCDPCQGALPDPSPSRVNPRTTRPE MLLRPESC PARQAPVCGDDGVTYENDCVMGRSGAARGLLLQKVRSGQCQGRDQCPEPC RFNAVCLSRGRPRCSCDRVTCDGAYRPVCAQDGRYDSDCWRRQAECRQRAIPSKH QGPCDQAPSPCLGVQCAFGATCAVKNGQAACECLQACSSLYDPVCGSDGVTYGSACEL EATACTLGREIQVADRCGQCRFGALCEAETGRCVCPSECVLAQPVCGSDGHTYTPSEC MLHVHACTHQISLHVASAGPCETCGDAVCAFGAVCSAGQCVCPRCEHPPPGPVCGSDG VTYGSACELREAACLQQTQIEEARAGPCEQAECGSGSGSGEDGDCEQLCRQRGGIW DEDSEDGPCVDFSCQSVPGSPVCGSDGVTYSTECCLKARCESQRGLYVAAQGACRG PTFAPLPVPAPLHCAQTPYGCCQDNITAAARGVGLAGCPSACQCNPHGSGYGGTCDPATG QCSCRPGVGGRLCDRCEPGFVNFGRGIVTDGRSGCTPCSCDPQGAVRDDCEQMTGLCSC KPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGARCVEESGSAHCVCML TCPEANATKVCSDGVTYGNECQLKTIACRQGLQISIQSLGPCQEA VAPSTHPTSA SV TVTTPGLLLSQALPAPPALPLAPSSSTAHSQTTPPPSSRPRTTASVPRTTVWPVLTVP PTAPSPAPSLVASAFGESGSTDGSSDEELSGDQEA SGGSGGLEPLEGSSVATPGPPV ERASCYNSALGCCSDGKTPSLDAEGSNCPATKVFQGVLELEGVEGQELFYTPEMADPK SELFGETARSIESTLDDLFRNSDVKKDFRSVRLRDLGPGKSVRAIVDVHFDPTTAFRA PDVARALLRQIQVSRRRSLGVRRPLQEHVRFMDFDWFFAFITGATSGAIAAGATARAT TASRLPSSAVTPRAPHPSHTSQPVAKTTAAPTRRPPTTAPSRVPGRRPPAPQPPKP CDSQPCFHGGTCQDWALGGGFTCSCPAGRGGAVCEKVLGAPVPAFEGRSFLAFTLRA YHTLRLALEFRALEPQGLLLYNGNARGKDFLALALLDGRVQLRFDTGSGFAVLTSAPV VEPGQWHRLELSRHWRRTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAAVA LERTFVGAGLRGCI RLLDVNNQRLELGIGPGAATRGSGVGECGDHPCLPNPCHGGAPC QNLEAGRFHCQCPPGRVGPCTADEKSPQPNPCHGAAPCRVLPEGGAQCECPLGREGT FCQTASGQDGSFPFLADFNFSHLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYN GQKTDGKGFVSLALRRLEFRYDLGKGA AVIRSREPVTLGAWTRVSLERNRKGAL RVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARA AAVSSGFDGAIQLVSLGGR QLLTPEHVL RQVDVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEK GLVEKSAGD VDTLAFDGRTFVEYLN AVTESEKALQSNHFELSLRTEATQGLVLWSGKA TERADYVALAIVDGHQLQSYNLG SQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGN EAPVTGSSPLGATQLD TDGALWLGGLPELPVGPALPKAYGTGFVGC LRDVVVGRHPLH LLEDAVTKPELRPCPTP		

Further analysis of the NOV40a protein yielded the following properties shown in Table 40B.

Table 40B. Protein Sequence Properties NOV40a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Likely cleavage site between residues 57 and 58

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 40C.

Table 40C. Geneseq Results for NOV40a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW26609	Human agrin - Homo sapiens, 492 aa. [WO9721811-A2, 19-JUN-1997]	1473..1931 22..492	458/471 (97%) 459/471 (97%)	0.0
AAB93754	Human protein sequence SEQ ID NO:13424 - Homo sapiens, 413 aa. [EP1074617-A2, 07-FEB-2001]	465..850 1..386	382/386 (98%) 385/386 (98%)	0.0
AAY73993	Human prostate tumor EST fragment derived protein #180 - Homo sapiens, 416 aa. [DE19820190-A1, 04-NOV-1999]	1516..1931 1..416	416/416 (100%) 416/416 (100%)	0.0
AAB31889	Amino acid sequence of a human protein - Homo sapiens, 4393 aa. [WO200105422-A2, 25-JAN-2001]	1237..1930 3639..4393	253/790 (32%) 352/790 (44%)	7e-90
ABB10233	Human cDNA SEQ ID NO: 541 - Homo sapiens, 432 aa. [WO200154474-A2, 02-AUG-2001]	1489..1928 3..429	142/449 (31%) 216/449 (47%)	2e-53

- In a BLAST search of public sequence databases, the NOV40a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 40D.

Table 40D. Public BLASTP Results for NOV40a				
Protein Accession Number	Protein/Organism/Length	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O00468	AGRIN PRECURSOR - Homo sapiens (Human), 2026 aa (fragment).	51..1931 137..2026	1843/1890 (97%) 1855/1890 (97%)	0.0
P25304	Agrin precursor - Rattus norvegicus (Rat), 1959 aa.	1..1931 1..1959	1561/1964 (79%) 1678/1964 (84%)	0.0
P31696	Agrin precursor - Gallus gallus (Chicken), 1955 aa.	51..1928 40..1952	1178/1931 (61%) 1416/1931 (73%)	0.0
Q90404	Agrin - Discopyge ommata (Electric ray), 1328 aa (fragment).	598..1929 1..1325	731/1353 (54%) 930/1353 (68%)	0.0

Q96IC1	UNKNOWN (PROTEIN FOR IMAGE:3544662) - Homo sapiens (Human), 488 aa (fragment).	1444..1931 1..488	488/488 (100%) 488/488 (100%)	0.0
--------	--	----------------------	----------------------------------	-----

PFam analysis predicts that the NOV40a protein contains the domains shown in the Table 40E.

Table 40E. Domain Analysis of NOV40a			
Pfam Domain	NOV40a Match Region	Identities/ Similarities for the Matched Region	Expect Value
thyroglobulin_1: domain 1 of 1	89..133	14/82 (17%) 31/82 (38%)	8.7
kazal: domain 1 of 9	89..133	24/61 (39%) 38/61 (62%)	3.6e-19
EGF: domain 1 of 11	133..176	13/47 (28%) 23/47 (49%)	23
kazal: domain 2 of 9	163..208	21/62 (34%) 33/62 (53%)	5.1e-13
kazal: domain 3 of 9	233..280	18/61 (30%) 33/61 (54%)	7.9e-12
EGF: domain 2 of 11	286..312	8/47 (17%) 19/47 (40%)	39
kazal: domain 4 of 9	307..352	21/61 (34%) 38/61 (62%)	4.1e-16
kazal: domain 5 of 9	381..426	23/62 (37%) 39/62 (63%)	1.7e-13
EB: domain 1 of 1	393..453	16/68 (24%) 35/68 (51%)	3.6
EGF: domain 3 of 11	423..453	9/47 (19%) 17/47 (36%)	1.3e+02
kazal: domain 6 of 9	441..485	19/61 (31%) 38/61 (62%)	1.5e-18
EGF: domain 4 of 11	493..518	10/47 (21%) 19/47 (40%)	99
kazal: domain 7 of 9	506..550	26/62 (42%) 37/62 (60%)	1.5e-17
kazal: domain 8 of 9	591..636	24/62 (39%) 40/62 (65%)	1.2e-16

EGF: domain 5 of 11	675..709	13/49 (27%) 23/49 (47%)	24
laminin_EGF: domain 1 of 2	679..730	28/61 (46%) 46/61 (75%)	1.2e-20
EGF: domain 6 of 11	735..763	10/49 (20%) 20/49 (41%)	18
laminin_EGF: domain 2 of 2	733..777	21/59 (36%) 37/59 (63%)	4e-11
EGF: domain 7 of 11	787..823	12/47 (26%) 22/47 (47%)	5.1
kazal: domain 9 of 9	809..855	25/62 (40%) 41/62 (66%)	5.3e-18
SEA: domain 1 of 1	1016..1138	39/132 (30%) 112/132 (85%)	1.4e-36
EGF: domain 8 of 11	1219..1252	16/47 (34%) 24/47 (51%)	0.00054
laminin_G: domain 1 of 3	1286..1417	70/162 (43%) 119/162 (73%)	3.1e-53
EGF: domain 9 of 11	1439..1471	16/47 (34%) 27/47 (57%)	5.1e-06
EGF: domain 10 of 11	1478..1510	16/47 (34%) 25/47 (53%)	0.0002
laminin_G: domain 2 of 3	1554..1685	70/161 (43%) 119/161 (74%)	6.6e-49
EGF: domain 11 of 11	1704..1738	14/47 (30%) 25/47 (53%)	2.3e-06
laminin_G: domain 3 of 3	1783..1914	59/161 (37%) 125/161 (78%)	1.7e-50

**EXAMPLE 41.**

The NOV41 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 41A.

Table 41A. NOV41 Sequence Analysis		
	SEQ ID NO: 181	770 bp
NOV41a, CG59895-01 DNA Sequence	ATGGGCAAAACACTGGACACAGACTGGATATAAAGACAGATGAGCTGGGGAGTGGAGC CCACTGCTAGAGAAAGACCCATCCCCAGCAACTGTGGAGGAGGCAGTGTCTGTCCCTTA CCAAGATGATGCTGCTGTTGCTGTGTCTGGGGTTGACCCTCGTCTGTGCCAGGAGGA AGAAAACATTTCAGGAGAGTGGTATTCGGTTCTCTTGGCCTCTGACTGCAGGGAAAAG ATAGAAGAAGATGGAAGCATGAGGGTTTTGTCAAACACATTGATTACCTGGGGAATT CTTCTCTGACTTTTAAATTGCATGAAATAAATGGAACTGTACTGAAATTAATTGGC TTGTAAACCAACAGAAAAGAACGGACTTAATGTCATTGACATACTTGAAACGGACTAT GATAATTATATATATTTTATAACAAGAATATCAAGAATGGGAAAACATTCCTAATGC	

	TGGAGCTCTATGGTAGAACACCGGATGTGAGCTCACAACTCAAGGAGAGGTTTGTGAA ATATTGTGAAGAACATGGGATTGATAAGGAAAACATATTTGACTTGACCAAAACAGAT CGCTGTCTCCAGGCCCGAGATGAGGGAGCAGCCTAGGACTCCGGGTTGGTGTATCTCTG ACACCGGTGGAGAGAGGGTGGCCCAGGGACCAGTGCCTTCCAAAAGCATTAGGGGTTT GCACCCAAAGATACCATAAAAATAATTTGGTAGGAAAGCTTGTGGGAAAATCTTGAA TCTGGAGTTGGAAGGT		
	ORF Start: ATG at 122	ORF Stop: TAG at 614	
	SEQ ID NO: 182	164 aa	MW at 18854.2kD
NOV41a, CG59895-01 Protein Sequence	MMLLLLCLGLTLVCAQEEENISGEWYSVLLASDCREKIEEDGSMRVFVKHIDYLGNS LTFKLHEINGNCTEINLACKPTEKNGLNVIDILETDYDNYIYFYNKNIKNGETFLMLE LYGRTPDVSSQLKERFVKYCEEHGIDKENIFDLTKTDRCLQARDEGAA		
	SEQ ID NO: 183	597 bp	
NOV41b, CG59895-02 DNA Sequence	GAGGAGGCAGTGTCTGCCCTTACCAAGATGATGCTGCTGTGTGCTGTCTGGGGTTGA CCCTCGTCTGTGCCAGGAGGAAGAAAACAATGATGCTGTGACAAGCAACTTCGATCT GTCAAAGATTTCAAGAGAGTGGTATTTCGGTTCTCTTGGCCTCTGACTGCAGGGAAG ATAGAAGAAGATGGAAGCATGAGGGTTTTTGTCAAACACATTGATTACCTGGGGAATT CTTCTCTGACTTTTAAATTGCATGAAATAGAAAATGGAACTGTACTGAAATTAATTT GGCTTGTAACCAACAGAAAAGAATTGTGTTGTTTCTCCACAGATAACGGACTTAAT GTCATTGACATACTTGAAACGGACTATGATAATTATATATATTTTTATAACAAGAATA TCAAGAATGGGGAACATTCTTAATGCTGGAGCTCTATGGTAGAACACCGGATGTGAG CTCACAACCTCAAGGAGAGGTTTGTGAAATATTGTGAAGAACATGGGATTGATAAGGAA AACATATTTGACTTGACCAAAGTTGGTAAGTCGGGGTTTCTGGTATTCTCTTCTTAA TTCCCATGTTACAGAAG		
	ORF Start: ATG at 28	ORF Stop: TAA at 577	
	SEQ ID NO: 184	183 aa	MW at 20803.3kD
NOV41b, CG59895-02 Protein Sequence	MMLLLLCLGLTLVCAQEEENNDAVTSNFDLSKISGEWYSVLLASDCREKIEEDGSMRV FVKHIDYLGNSLTFKLHEIENGNCIEINLACKPTEKNVSVSTDNGLNVIDILETDY DNYIYFYNKNIKNGETFLMLELYGRTPDVSSQLKERFVKYCEEHGIDKENIFDLTKVG KSGFLVFSS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 41B.

Table 41B. Comparison of NOV41a against NOV41b and NOV41c.		
Protein Sequence	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV41b	13..154	124/163 (76%)
	13..175	125/163 (76%)

Further analysis of the NOV41a protein yielded the following properties shown in Table 41C.

Table 41C. Protein Sequence Properties NOV41a	
PSort analysis:	0.4180 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Likely cleavage site between residues 16 and 17

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 41D.

<b>Table 41D. Geneseq Results for NOV41a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV41a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAG68142	Rat TRDH-110 protein sequence SEQ ID NO:10 - Rattus norvegicus, 181 aa. [WO200173022-A1, 04-OCT-2001]	1..159 3..180	94/179 (52%) 121/179 (67%)	5e-45
AAU29121	Human PRO polypeptide sequence #98 - Homo sapiens, 180 aa. [WO200168848-A2, 20-SEP-2001]	2..160 3..180	87/179 (48%) 117/179 (64%)	2e-40
AAB65225	Human PRO1054 (UNQ519) protein sequence SEQ ID NO:256 - Homo sapiens, 180 aa. [WO200073454-A1, 07-DEC-2000]	2..160 3..180	87/179 (48%) 117/179 (64%)	2e-40
AAY66702	Membrane-bound protein PRO1054 - Homo sapiens, 180 aa. [WO9963088-A2, 09-DEC-1999]	2..160 3..180	87/179 (48%) 117/179 (64%)	2e-40
AAY25674	Horse allergen 1575778 Equ c 1 protein fragment - Equus sp, 187 aa. [WO9934826-A1, 15-JUL-1999]	1..164 1..185	92/185 (49%) 110/185 (58%)	2e-39

- In a BLAST search of public sequence databases, the NOV41a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 41E.

<b>Table 41E. Public BLASTP Results for NOV41a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV41a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P11590	Major urinary protein 4 precursor (MUP 4) - Mus musculus (Mouse), 178 aa.	2..160 1..178	104/179 (58%) 124/179 (69%)	1e-47
Q63213	ALPHA-2U GLOBULIN (RAT SALIVARY GLAND (ALPHA)2(MU) GLOBULIN, TYPE 1) - Rattus norvegicus (Rat), 181 aa.	1..159 3..180	98/179 (54%) 124/179 (68%)	2e-47
S05440	alpha-2u-globulin precursor - rat, 179 aa.	1..158 3..179	97/178 (54%) 123/178 (68%)	6e-47



Q9JJI1	ALPHA-2U GLOBULIN - Rattus norvegicus (Rat), 181 aa.	1..159 3..180	96/179 (53%) 123/179 (68%)	4e-46
Q9JJI3	ALPHA-2U GLOBULIN - Rattus norvegicus (Rat), 181 aa.	1..159 3..180	96/179 (53%) 122/179 (67%)	2e-45

PFam analysis predicts that the NOV41a protein contains the domains shown in the Table 41F.

Table 41F. Domain Analysis of NOV41a			
Pfam Domain	NOV41a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lipocalin: domain 1 of 1	20..155	50/157 (32%) 111/157 (71%)	3e-32

#### EXAMPLE 42.

The NOV42 clone was analyzed, and the nucleotide and predicted polypeptide  
5 sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis		
	SEQ ID NO: 185	4205 bp
NOV42a, CG59889-01 DNA Sequence	ATTAATGAATATAAAATTATTATGTACTACACAATTAGTAGAAAGCATATTTAGAGA CACACCTGCCGCAAAATACTCAGTCAAGGGAAGGGGCGGGTCCGAATCCAGGGGCGAC GCCGCCGCTCCGCCAGTGTCCCGGGCGTCCCGCCGCTCCTAAGCGCTGGAGCGC GAGGATCGCTCCACTGCCTCCAGCCTGGGCAACAGAGCGAGACTCTGTCTCAAAAAA AAAAAAGAAGTAAAAATAATTATGCAGTATGTTTAGACATTTAATATTGTTTGTAT TTCATTTTTTCTTCCCTTAAAAACACCCCTTGGGGAGACTTCGGCTGTGGGTGCCCT GACCAGAGCCCTGAGTTGCAACCCTGGAACCCTGGCCATGACCAAGACCACCATGTGC ATATCGGCCAGGCAAGACACTGCTGCTCACCTCTCTGCCACGGTCTATTCCATCCA CATCTCAGAGGGAGGCAAGCTGGTCATTAAAGACCACGACGAGCCGATTGTTTTGCGA ACCCGGCACATCCTGATTGACAACGGAGGAGAGCTGCATGCTGGGAGTGCCCTCTGCC CTTTCCAGGGCAATTTACCATCATTGTGTATGAAGGGCTGATGAAGGTATTACAGCC GGATCCTTACTATGGTCTGAAGTACATTGGGGTGGTAAAGGAGGCGCTCTTGAGTTG CATGGACAGAAAAAGCTCTCCTGGACATTTCTGAACAAGACCCTTCACCCAGGTGGCA TGGCAGAAGGAGGCTATTTTTTTGAAAGGAGCTGGGGCCACCGTGGAGTTATTGTTCA TGTCATCGACCCCAAATCAGGCACAGTCATCCATTCTGACCGGTTTGACACCTATAGA TCCAAGAAAGAGAGTGAACGTCTGGTCCAGTATTTGAACGCGGTGCCGATGGCAGGA TCCTTTCTGTTGCAGTGAATGATGAAGGTTCTCGAAATCTGGATGACATGGCCAGGAA GGCGATGACCAAATTTGGGAAGCAAACACTTCTCTGCACCTTGGATTAGGGTGGAGTGG ACGGAGTGGTTCGATCATGATAAAGTATCTCAGACTAAAGGTGGGAGAAAATTCAG ACCTCTGGAAAGCTCACCCAGGAAAAATATGCAATCGTCCCATTTGATATACAGCAGGC CACTACAATGGATGGAGTTAACCTCAGCACCGAGGTTGTCTACAAAAAAGGCCAGGAT TATAGGTTTGCTTGCTACGACCGGGGAGAGCCTGCCGAGCTACCGTGACGGTTC TCTGTGGGAAGCCTGTGAGGCCCAAACCTCACAGTCACCATTGACACCAATGTGAACAG CACCATTTGAACCTGGAGGATAATGTACAGTCATGGAACCTGGAGATACCCCTGGTC ATTGCCAGTACTGATTACTCCATGTACAGGCAGAAGAGTTCAGGTGCTTCCCTGCA GATCCTGCGCCCCCAACAGGTCAAAGTGGCAGGGAAACCAATGTACCTGCACATCGG GGAGGAGATAGACGCGTGGACATGCGGGCGGAGGTTGGGCTTCTGAGCGGAACATC ATAGTGATGGGGGAGATGGAGGACAAATGCTACCCCTACAGAAACCACATCTGCAATT TCTTTGACTTCGATACCTTTGGGGGCCACATCAAGTTTGCTCTGGGATTTAAGGCAGC ACACCTTGAGGGCACGGAGCTGAAGCATATGGGACAGCAGCTGGTGGGTGAGTACCCG ATTCACTTCCACCTGGCCGGTGATGTAGACGAAAGGGAGGTTATGACCCACCCACAT	

	<p>ACATCAGGGACCTCTCCATCCATCATACATTCTCTCGCTGCGTCACAGTCCATGGCTC  CAATGGCTTGTGTGATCAAGGACGTTGTGGGCTATAACTCTTTGGGCCACTGCTTCTTC  ACGGAAGATGGGCCGGAGGAACGCAACACTTTTGACCACTGTCTTGGCCTCCTTGTCA  AGTCTGGAACCTCCTCCCTCGGACCGTGACAGCAAGATGTGCAAGATGATCACAGA  GGACTCCTACCCAGGGTACATCCCCAAGCCCAGGCAAGACTGCAATGCTGTGTCCACC  TTCTGGATGGCCAATCCCAACAACAACCTCATCAACTGTGCCGTGCAGGATCTGAGG  AAACTGGATTTTGGTTTATTTTCCACAGTACCAACGGGCCCTCCGTGGGAATGTA  CTCCCCAGGTTATTAGAGCACATTCCACTGGGAAAATCTATAACAACCGAGCACAT  TCCAACACCGGGCTGGCATGATCATAGACAACGGAGTCAAAACCACCGAGGCTCTG  CCAAGGACAAGCGGCCCTTCTCTCAATCATCTCTGCCAGATACAGCCCTACCAGGA  CGCCGACCCGCTGAAGCCCCGGGAGCCGGCCATCATCAGACACTTATTGCCTACAAG  AACCAGGACCACGGGGCTGGCTGCGCGCGGGGATGTGTGGCTGGACAGTGCCTGGT  TTGCTGACAATGGCATTGGCCTGACCTGGCCAGTGGTGGAAACCTTCCCGTATGACGA  CGGCTCCAAGCAAGAGATAAAGAACAGCTTGTGTGGCGAGAGTGGCAACGTGGGG  ACGGAATGATGGACAATAGGATCTGGGGCCCTGGCGGCTTGGACCATAGCGGAAGGA  CCCTCCTATAGGCCAGAAATTTCCAATTAGAGGAATTAGTTATATGATGGCCCCAT  CAACATCCAAACTGCACTTTCCGAAAGTTTGTGGCCCTGGAGGGCCGGCACACCAGC  GCCCTGGCCTTCGCCCTGAATAATGCCTGGCAGAGCTGCCCAATACCAAGTGCACCG  GCATTGCCTTTGAGGACGTTCCGATTACTTCCAGAGTGTCTTCGGAGAGCCTGGGCC  CTGGTTCAACCAGCTGGACATGGATGGGGATAAGACATCTGTGTTCATGACGTCGAC  GGCTCCGTGTCCGAGTACCCTGGCTCCTACCTACGAAGAATGACAACTGGCTGGTCC  GGCACCAGACTGCATCAATGTTCCCGACTGGAGAGGGGCCATTTGCAGTGGGTGCTA  TGCACAGATGTACATTCAAGCCTACAAGACCAGTAACCTGCGAATGAAGATCATCAAG  AATGACTTCCCCAGCCACCCTCTTACCTGGAGGGGGCGCTCACCAGGAGCACCATT  ACCAGCAATACCAACCGGTTGTACCCTGCAGAAGGGCTACACCATCCACTGGGACCA  GACGGCCCCCGCAACTCGCCATCTGGCTCATCAACTTCAACAAGGGCGACTGGATC  CGAGTGGGGCTCTGCTACCCGCGAGGCACCACATTCTCCATCCTCTCGGATGTTACA  ATCGCCTGCTGAAGCAAACGTCGAAGACGGGCGTCTTCGTGAGGACCTTGCAGATGGA  CAAAGTGGAGCAGAGCTACCTGGCAGGAGCCACTACTACTGGGACGAGACTCAGGG  CTGTTGTTCTGAAGCTGAAAGCTCAGAACGAGAGAGAGAAGTTGCTTTCTGCTCCA  TGAAAGGCTGTGAGAGGATAAAGATTAAAGCTCTGATTCCAAAGAACGCAGGCGTCAG  TGACTGCACAGCCACAGCTTACCCCAAGTTCACCGAGAGGGCTGTCTGATAGACGTGCCG  ATGCCCAAGAAGCTCTTTGGTTCTCAGCTGAAAACAAAGGACCATTTCTTGGAGGTGA  AGATGGAGAGTTCCAAGCAGCACTTCTTCCACCTCTGGAACGAGTTCGTTACATTGA  AGTGGATGGGAAGAAGTACCCCAAGTTCGGAGGATGGCATCCAGGTGGTGGTATTGAC  GGGAACCAAGGGCGCGTGGTGAGCCACACGAGCTTCAGGAATCCATTCTGCAAGGCA  TACCATGGCAGCTTTTCAACTATGTGGCGACCATCCCTGACAATTCCATAGTGCTTAT  GGCATCAAAGGGAAGATACGTCTCCAGAGGCCCATGGACCAGAGTGTGGAAGAGCTT  GGGGCAGACAGGGGTCTCAAGTTGAAAGAGCAAATGGCATTCTGTTGGCTTCAAAGGCA  GCTTCCGGCCCCATCTGGGTGACACTGGACACTGAGGATCACAAGGCCAAATCTTCCA  AGTTGTGCCCATCCCTGTGGTGAAGAAGAAGATTGTGAGGACAGCTGCCGCCGGT  GCCACCTCGTGGTAGACTATGACGGTGAC</p>
	<p>ORF Start: ATG at 22      ORF Stop: TGA at 4156</p>
	<p>SEQ ID NO: 186      1378 aa      MW at 155014.9kD</p>
<p>NOV42a, CG59889-01 Protein Sequence</p>	<p>MYTISRKHILETHLPQNTQSREGAGPNPGATPPPPVPRASRLTKRLEREDRSTAL  QPGQQSETLSQKKRKNYAVCLDILIFVLISFFLPLKTPLETSAAGCPDQSPQLQ  PWNPGHDQDHHVHIGQKTLTLLTSSATVYSIHISEGGLVKIDHDEPIVLRTRHILID  NGGELHAGSALCPFQGNFTIILYGRADEGIQDPYGLKYIGVGKGGALELHGQKKLS  WTFNLKTLHPGMAEGGYFFERSWGRGVIHVVIDPKSGTVIHSDFDTRYRKKESER  LVQYLNAPVDGRILSVAVNDEGSRNLDDMARKAMTKLGSKHFLHLGFRFEWTEWFDHD  KVSQTKGGEKISDLWKAHPGKICNRPIDIQATTMDGVNLSTEVVYKKGQDYRFACYD  RGRACRSYRVRFLCGKPVPRKLTVTIDTNVNSTILNLEDNVQSWKPGDTLVIASDYS  MYQAEFFQVLPSCRAPNQVKVAGKPMYLIHIGEEIDGVDMAEVGLLSRNIIVMGEME  DKCYPYRNHICNFFDFTFGGHIKFALGFKAHLEGTELKHMGGQLVGQYPIHFHLAG  DVDERGGYDPPTYIRDLSIHHTFSRCVTVHGSNGLLIKDVVGYNLSLGHCFFTEDGPEE  RNTFDHCLGLLVKSGTLLPSDRDSKMCKMITEDSYPGYIPKPRQDCNAVSTFWMANPN  NNLINCAAAGSEETGFWFIFHHVPTGPSVGMSPGYSEHILGKFYNNRAHSNYRAGM  IIDNGVKTTEASAKDKRFLSIISARYSPHQDADPLKPREPAIRHFYIAYKNQDHGAW  LRGGDVWLDSCRFA DNIGILTLASGGTFPYDDGSKQEIKNLSLVGESGNVGTMMMDNR</p>

	IWGPGGLDHSGRTP LPIGQNFPIRGIQLYDGPINIQNCTFRKFVALEGRHTSALAFRLN NAWQSCPHNNVTGIAFEDVPITSRVFFGEPGPFNQLDMDGDKTSVFHDVDGVSVEYP GSYLTQNDNLVLRHPDCINVPDWRGAICSGCYAQMYIQAYKTSNLRMKI IKNDPFSHP LYLEGALTRSTHYQQYQPVVTLQKGYTIHWDQTAPAE LAIWLINFNKGDWIRVGLCYP RGTTFSILSDVHNRLLKQTSKTVFVRTLQMDKVEQSYPRSHYYWDEDSGLLFLKLK AQNEREKFAFCSMKGCERIKIKALIPKNAGVSDCTATAYPKFTERAVVDVPM PKLFG SQLKTKDHFLEVKMESSKHFFHLW NDFAYIEVDGKKYPSSEDGIQVVVIDGNQGRVV SHTSFRNSILQGI PWQLFNYVATIPDNSIVLMASKGRYVSRGPWTRVLEKL GADRGLK LKEQMAFVGFKGSFRPIWVTLDTEDHKAKIFQVVP I PVVKKKKL	
	SEQ ID NO: 187	7233 bp
NOV42b, CG59889-02 DNA Sequence	GAGCTAGCGCTCAAGCAGAGCCAGCGCGGTGCTATCGGACAGAGCCTGGCGAGCGCA AGCGGCGCGGGAGCCAGCGGGGCTGAGCGCGGCCAGGGTCTGAACCCAGATTTCCTCA GACTAGCTACCACTCCGCTTGCCACGCCCCGGGAGCTCGCGCGCCTGGCGGTACG GACCAGACGTCCGGGGCGCTGCGCTCCTGGCCCGCAGGCGTGACACTGTCTCGGCT ACAGACCCAGAGGGAGCACACTGCCAGGATGGGAGCTGCTGGGAGGCAGGACTTCCTC TTCAAGGCCATGCTGACCATCAGCTGGCTCACTCTGACCTGCTTCCTGGGGCCACAT CCACAGTGGCTGCTGGGTGCCCTGACCAGAGCCTGAGTGGCAACCTGGAACCTGG CCATGACCAAGACCACCATGTGCATATCGGCCAGGGCAAGACACTGCTGCTCACCTCT TCTGCCACGGTCTATTCCATCCACATCTCAGAGGGAGGCAAGCTGGTCATTAAAGACC ACGACGAGCCGATTGTTTTGCGAACC CGGCACATCCTGATTGACAACGGAGGAGAGCT GCATGCTGGGAGTGCCCTCTGCCCTTTCCAGGGCAATTTACCATCATTTTGTATGGA AGGGCTGATGAAGGTATTAGCCGGATCCTTACTATGGTCTGAAGTACATTGGGGTTG GTAAAGGAGGCGCTCTTGAGTTGCATGGACAGAAAAAGCTCTCTGGACATTTCTGAA CAAGACCCTTACCCAGGTGGCATGGCAGAAGGAGGCTATTTTTTTGAAAGGAGCTGG GGCCACCGTGGAGTTATTGTTTCATGTCATCGACCCCAAATCAGGCACAGTCATCCATT CTGACCGGTTTGACACCTATAGATCCAAGAAAGAGAGTGAACGTCTGGTCCAGTATTT GAACGCGGTGCCCGATGGCAGGATCCTTTCTGTTGCAGTGAATGATGAAGGTTCTCGA AATCTGGATGACATGGCCAGGAAGCGATGACCAAATTGGGAAGCAAACTTCTCTGC ACCTTGGATTTAGACACCTTGGAGTTTCTAACTGTGAAAGGAAATCCATCATCTTC AGTGGAAGACCATATTGAATATCATGGACATCGAGGCTCTGCTGCTGCCCGGTATTC AAATTGTTCCAGACAGAGCATGGCGAATATTTCAATGTTTCTTTGTCAGTGAGTGGG TTCAAGACGTGGAGTGGACGGAGTGGTTCGATCATGATAAAGTATCTCAGACTAAAGG TGGGGAGAAAATTTAGACCTCTGGAAAGCTCACCAGGAAAAATATGCAATCGTCCC ATTGATATACAGGCCACTACAATGGATGGAGTTAACTCAGCACCGAGGTTGTCTACA AAAAAGGCCAGGATTATAGGTTTGCTTGCTACGACCGGGGAGAGCCTGCCGGAGCTA CCGTGTACGGTTCCTCTGTGGGAAGCCTGTGAGGCCCAAACCTCACAGTCACCATTGAC ACCAATGTGAACAGCACCATTTCTGAACCTGGAGGATAATGTACAGTCATGGAACCTG GAGATACCTGGTCATTGCCAGTACTGATTACTCCATGTACCAGGCAGAAGATTCCA GGTGCTTCCCCTGCAGATCCTGCGCCCCCAACCAGGTCAAAGTGGCAGGCAACCAATG TACCTGCACATCGGGGAGGAGATAGACGGCGTGGACATGCGGGCGGAGGTTGGGCTTC TGAGCCGGAACATCATAGTGATGGGGGAGATGGAGGACAAATGCTACCCCTACAGAAA CCACATCTGCAATTTCTTTGACTTCGATACTTTGGGGGCCACATCAAGTTTGCTCTG GGATTTAAGGCAGCACACTTGGAGGGCAGGAGCTGAAGCATATGGGACAGCAGCTGG TGGGTCAGTACCGGATTCATTTCCACCTGGCCGGTGATGTAGACGAAAGGGAGGTTA TGACCCACCCACATACATCAGGGACCTCTCCATCCATCATACATTCTCTCGCTGCGTC ACAGTCCATGGCTCCAATGGCTTGTGATCAAGGACGTTGTGGGCTATAACTCTTTGG GCCACTGCTTCTTCACGGAAGATGGGCCGGAGGAACGCAACACTTTTGACCACTGTCT TGGCCTCCTTGTCAAGTCTGGAACCTCCTCCCTCGGACCGTGACAGCAAGATGTGC AAGATGATCACAGGAGTCTTACCCAGGTACATCCCCAAGCCAGGCAAGACTGCA ATGCTGTGTCCACCTTCTGGATGGCCAATCCCAACAACACTCATCAACTGTGCCG TGCAGGATCTGAGGAACTGGATTTTGGTTTATTTTACCACGTACCAACGGGCCCC TCCGTGGGAATGTACTCCCGAGTTATTTCAGAGCACATTCCACTGGGAAAATTTCTATA ACAACCGAGCACATTCCAACCTACCGGCTGGCATGATCATAGACAACGGAGTCAAAAC CACCAGGCCTCTGCCAAGGACAAGCGGCCGTTCTCTCAATCATCTCTGCCAGATAC AGCCCTCACCAGGACGCCGACCCGCTGAAGCCCCGGGAGCCGCGCATCATCAGACACT TCATTGCCTACAAGAACAGGACCACGGGGCCTGGCTGCGCGGCGGGGATGTGTGGCT GGACAGCTGCCGTTTGTGACAATGGCATTGGCTGACCTGGCCAGTGGTGGAAACC TTCCCGTATGACGACGGCTCAAGCAAGAGATAAAGAACAGCTTGTTTGTTGGCGAGA GTGGCAACGTGGGGACGGAATGATGGACAATAGGATCTGGGGCCCTGGCGGCTTGA CCATAGCGGAAGGACCTCCCTATAGGCCAGAAATTTCCAATTAGAGGAATTCAGTTA	

TATGATGGCCCCATCAACATCCAAACTGCACCTTCCGAAAGTTGTGGCCCTGGAGG  
 GCCGGCACACGACGCGCCTGGCCTTCCGCTGAATAATGCCTGGCAGAGCTGCCCCA  
 TAACAACGTGACCGGCAITGCCTTTGAGGACGTTCCGATTACTTCCAGAGTGTTCTTC  
 GGAGAGCCTGGGCCCTGGTTCAACCAGCTGGACATGGATGGGGATAAGACATCTGTGT  
 TCCATGACGTCGACGGCTCCGTGTCCGAGTACCTGGCTCCTACCTCACGAAGATGA  
 CAACTGGCTGGTCCGGCACCCAGACTGCATCAATGTTCCCGACTGGAGAGGGGCCATT  
 TGCAGTGGGTGCTATGCACAGATGTACATTCAAGCCTACAAGACCAGTAACCTGCGAA  
 TGAAGATCATCAAGAATGACTTCCCCAGCCACCCTCTTTACCTGGAGGGGGCGTCCAC  
 CAGGAGCACCCATTACCAGCAATACCAACCGGTTGTACCCTGCAGAAGGGCTACACC  
 ATCCACTGGGACCAGACGGCCCCCGCGAACTCGCCATCTGGCTCATCAACTCAACA  
 AGGGCGACTGGATCCGAGTGGGGCTCTGCTACCCGCGAGGCCACCATTTCTCCATCCT  
 CTCGGATGTTACAATCGCTGTGAAGCAAACGTCCAAGACGGGCGTCTTCGTGAGG  
 ACCTTGAGATGGACAAAGTGGAGCAGAGCTACCCTGGCAGGAGCCACTACTACTGGG  
 ACGAGGACTCAGGGCTGTTGTTCTGAAAGCTGAAAGCTCAGAACGAGAGAGAGAAGTT  
 TGCTTTCTGCTCCATGAAAGGCTGTGAGAGGATAAAGATTAAAGCTCTGATTCCAAAG  
 AACGCAGGCGTCAGTGACTGCACAGCCACAGCTTACCCCAAGTTACCGAGAGGGCTG  
 TCGTAGACGTGCCGATGCCAAGAAGCTCTTTGGTTCTCAGCTGAAAACAAAGGACCA  
 TTTCTGGAGGTGAAGATGGAGAGTTCCAAGCAGCACTTCTCCACCTCTGGAACGAC  
 TTCGCTTACATTGAAGTGGATGGGAAGAAGTACCCAGTTCGGAGGATGGCATCCAGG  
 TGGTGGTGATTGACGGGAACCAAGGGCGCGTGGTGAGCCACACGAGCTTCAGGAATC  
 CATTCTGCAAGGCATACCATGGCAGCTTTTCAACTATGTGGCGACCATCCCTGACAAT  
 TCCATAGTGCTTATGGCATCAAAGGGAAGATACGTCTCCAGAGGCCCATGGACCAGAG  
 TGCTGGAAAAGCTTGGGCGAGACAGGGGTCTCAAGTTGAAAGAGCAAATGGCATTCTGT  
 TGGCTTCAAAGGCAGCTTCCGGCCCATCTGGGTGACACTGGACACTGAGGATCACAAA  
 GCCAAAATCTTCAAGTTGTGCCCATCCCTGTGGTGAAGAAGAAGTGTGTGAGGAC  
 AGCTGCCGCCCGGTGCCACCTCGTGGTAGACTATGACGGTGACTCTTGGCAGCAGACC  
 AGTGGGGGATGGCTGGGTCCCCAGCCCCCTGCCAGCAGCTGCCTGGGAAGGCCGTGTT  
 TCAGCCCTGATGGGCCAAGGGAAGGCTATCAGAGACCCTGGTGCTGCCACCTGCCCT  
 ACTCAAGTGTCTACCTGGAGCCCTGGGGCGGTGCTGGCCAATGCTGGAACATTAC  
 TTTCTGCGACCTCTTGGGTGCTTCTCTCTATCTGTGCCTCTTCACTGGGGGTTGG  
 GGACCATATCAGGAGACCTGGGTTGTGCTGACAGCAAAGATCCACTCTGGCAGGAGCC  
 CTGACCCAGCTAGGAGGTAGTCTGGAGGGCTGGTCATTACAGATCCCCATGGTCTTC  
 AGCAGACAAGTGAGGGTGGTAAATGTAGGAGAAAGAGCCTTGGCCTTAAAGAAATCTT  
 TACTCTGTGAAGCAAGAGCCAACCTCACAGGATTAGGAGCTGGGGTAGAAGTGGCTAT  
 CCTTGGGGAAGAGGCAAGCCCTGCCCTCTGGCGGTGTCCACCTTTTCAAGAGACTTTGAG  
 TGGCAGGTTTGGACTTGGACTAGATGACTCTCAAAGGCCCTTTTAGTTCTGAGATTCC  
 AGAAATCTGCTGCATTTACATGGTACCTGGAACCCAACAGTTTCATGGATATCCACTG  
 ATATCCATGATGCTGGGTGCCCCAGCGCACACGGGATGGAGAGGTGAGAACTAATGCC  
 TAGCTTGAGGGGTCTGCAGTCCAGTAGGGCAGGCAGTCAGTCCATGTGCAGTGCAT  
 GCCAGGTGGAGAAATCACAGAGAGGTAAATGGAGGCCAGTGCCTTACAGAGGGAG  
 GCTCAGGAAGGCTTCTTGCTTACAGGAATGAAGCTGGGGGCATTTTGTGTTGGGGGAG  
 ATGAGGCAGCCTCTGGAATGGCTCAGGGATTACGCCCTCCCTGCCGTGCTGTGAA  
 GCTGGTGACTACGGGGTCCGCCCTTGTCTCACGTCTCTCTGCCCACTCATGATGGAGA  
 AGTGTGGTCAGAGGGGAGCAATGGGCTTTGTGCTTATGAGCACAGAGGAATTCAGTC  
 CCCAGGCAGCCCTGCCTCTGACTCCAAGAGGGTGAAGTCCACAGAAGTGAGCTCCTGC  
 CTTAGGGCCTCATTGTCTTTCATCCAGGGAAGTGAAGCACAGGGGGCTCCAGGAGAC  
 CCTAGATGTGCTCGTACTCCCTCGGCCTGGGATTTTCAAGCTGGAATATAGAAAATA  
 TCTAGCCCAAAGCCTTCATTTTAACAGATGGGGAAAGTGAGCCCCAAGATGGGAAAG  
 AACCACACAGCTAAGGGAGGGCCTGGGGAGCCCCACCCTAGCCCTTGTGCCACACCA  
 CATTGCCTCAACAACCGGCCCCAGAGTGCCAGGCACTCCTGAGGTAGCTTCTGGAAA  
 TGGGGACAAGTCCCTCGAAGGAAAGGAAATGACTAGAGTAGAATGACAGCTAGCAGA  
 TCTCTTCCCTCCTGCTCCAGCGCACACAAACCGCCCTCCCTTGGTGTGGCGGTC  
 CCTGTGGCCTTCACTTTGTTCACCTGTCAGCCAGCCTGGGTGCACAGTAGCTGC  
 AACTCCCCATTGGTGCTACCTGGCTCTCTGTCTCTGCAGCTCTACAGGTTAGGCCCA  
 GCAGAGGGAGTAGGGCTCGCCATGTTTCTGGTGAGCCAATTTGGCTGATCTTGGGTGT  
 CTGAACAGCTATTGGGTCCACCCAGTCCCTTTCAGCTGCTGCTTAATGCCCTGCTCT  
 CTCCCTGGCCACCTTATAGAGAGCCCAAAGAGCTCCTGTAAGAGGGAGAACTCTATC  
 TGTGGTTTATAAGCTTGACGAGGACCAGAGTCTCCCTGGGTCTTGTGATGAACCTAC  
 ATTTATCCCTTTCCTGCCCAACCAAACTCTTTCCTTCAAAGAGGGCCTGCCCTGG  
 CTCCCTCCACCCAACTGCACCCATGAGACTCGGTCCAAGAGTCCATTCCCCAGGTGGG  
 AGCCAACTGTCAAGGAGTCTTCCACCAACATCTTTCAGCTGCTGGGAGGTGACC

	ATAGGGCTCTGCTTTTAAAGATATGGCTGCTTCAAAGGCCAGAGTCACAGGAAGGACT TCTTCCAGGGAGATTAGTGGTGATGGAGAGGAGAGTTAAAATGACCTCATGTCCTTCT TGTCACGGTTTTGTTGAGTTTTCACTCTTCTAATGCAAGGGTCTCACACTGTGAACC ACTTAGGATGTGATCACTTTCAGGTGGCCAGGAATGTTGAATGTCTTTGGCTCAGTTC ATTTAAAAAGATATCTATTTGAAAGTTCTCAGAGTTGTACATATGTTTCACAGTACA GGATCTGTACATAAAAGTTTCTTTCTAAACCATTACCAGAGGCCAATATCTAGGCA TTTTCTTGGTAGCACAAATTTTCTTATTGCTTAGAAAATTGTCCTCCTTGTTATTCT GTTTGTAAGACTTAAGTGAGTTAGGTCTTTAAGGAAAGCAACGCTCCTCTGAAATGCT TGCTTTTTTTCTGTTGCCGAAATAGCTGGTCTTTTTTCGGGAGTTAGATGTATAGAGT GTTTGTATGTAAACATTTCTTGTAGGCATCACCATGAACAAAGATATATTTTCTATTT ATTTATATATGTGCATTTCAAGAAGTCACTGTGAGAGAAATAACAATTTGTCTTAAA TGTCATGATTGGAGATGTCCTTTGCATTGCTTGAAGGGGTGTACCTAGAGCCAAGGA AATTGGCTCTGTTTGGAAAAATTTTGTCTGTTATTATAGTAAACATACAAAGGATGTC CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 261	ORF Stop: TGA at 4344	
	SEQ ID NO: 188	1361 aa	MW at 152996.4kD
NOV42b, CG59889-02 Protein Sequence	MGAAGRQDFLFKAMLTISWLTLCFPGATSTVAAGCPDQSPQLPWNPGHDQDHHVHI GQGKTLTSSATVYSIHISEGGKLVKDHDEPIVLRTRHLLIDNGGELHAGSALCPF QGNFTIILYGRADEGIQDPYPYGLKYIGVGKGALELHGQKLSWTFLNKTLHPGGMA EGGYFFERSWGHGRVIVHVIDPKSGTVIHSRDFDTRSKKESERLVQYLNAPVDRIL SVAVNDEGSRNLDDMARKAMTKLGSKHFLHLGFRHPWSFLTVMGNPSSSVEDHIEYHG HRGSAARVFKLFQTEHGEYFNVSLSEWVQDVEWTEWFDHDKVSQTKGGEKISDLWK AHPGKICNRPIDIQATMDGVNLSTEVVYKKGQDYRFACYDRGRACRSYRVRFLCCKP VRPKLTVTIDTNVNSTILNLEDNVQSWKPGDTLVIASTDYSMYQAEFFQVLPSCRSCAP NOVKVAGKPMYLHIGEEIDGVDMAEVLGSLRNIIVMGEMEDKCPYRNHICNFFDFD TFGGHIKFALGFKAHLEGTELKHMGGQLVGGYPIHFHLAGVDVERGGYDPTTYIRDL SIHHTFSRCVTVHGSNGLLIKDVVGYNSLGHCFFTEDGPEERNTFDHCLGLLVKSGTL LPSDRDSKMKMITEDSYPGYIPKPRQDCNAVSTFWMANPNNNLINCAAAGSEETGFW FIFHHVPTGPSVGMSPGYSEHIFLGKFYNNRAHSNRYRAGMIIDNGVKTTTEASAKDKR PFLSIISARYSPHQDADPLKPREPAIRHFIAIKNQDHGAWLRGGDVWLDSCRFDNG IGLTLASGGTFPYDDGSKQEIKNSLFVGESGNVGTEEMDNRIWGPGLDHSRGTLPIG QNFPIRGIQLYDGPINIQNCTFRKFVALEGRHTSALAFRLNNAWQSCPHNNVTGIAFE DVPITSRVFFGEPGFWNQDMDGDKTSVFHDVDGSVSEYPGSYLTKNDNWLVHRPDC INVDPWRGAICSGCYAQMYIQAYKTSNLRMKIKINDFPSPHPLYLEGALTRSTHYQQYQ PVVTLQKGYTIHWDQTAPELAIWLINFNKGDWIRVGLCYPRGTTFSILSDVHNRLLK QTSKTGVFVRTLQMDKVEQSYPRSHYYWDEDSGLLFLKLKAQNEREKFAFCSMKGCE RIKIKALIPKNAGVSDCTATAYPKFTERAVVDVPMFKKLFQSLKTKDHFLEVKMESS KQHFFHLWDFAYIEVDGKKYPSSEDDGIQVVVIDGNQGRVVSHTSFRNSILQGIWQQL FNYVATIPDNSIVLMAKGRYVSRGPWTRVLEKLGA DRGLKLKEQMAFVGFGKGSFRPI WVTLDTEDEHKAKIFQVVPIPVVKKKKL		
	SEQ ID NO: 189	3864 bp	
NOV42c, CG59889-04 DNA Sequence	GTGCCCTGACCAGAGCCCTGAGTTGCAACCCTGGAACCCTGGCCATGACCAAGACCAC CATGTGCATATCGGCCAGGCAAGACTGCTGCTCACCTCTTCTGCCACGGTCTATT CCATCCACATCTCAGAGGGAGGCAAGCTGGTCATTAAAGACCACGACGAGCCGATTGT TTTGCGAACC CGGCACATCCTGATTGACAACGAGGAGAGCTGCATGCTGGGAGTGCC CTCTGCCCTTTCCAGGGCAATTTACCATCATTTTGTATGGAAGGGCTGATGAAGGTA TTCAGCCGGATCCTTACTATGGTCTGAAGTACATTGGGGTTGGTAAAGGAGGCCCTCT TGAGTTGCATGGACAGAAAAAGCTCTCCTGGACATTTCTGAACAAGACCCTTCACCCA GGTGGCATGGCAGAAGGAGGCTATTTTTTTGAAAGGAGCTGGGGCCACCGTGGAGTTA TTGTTTCATGTCATCGACCCCAATCAGGCACAGTCATCCATTCTGACCGGTTTGACAC CTATAGATCCAAGAAAGAGAGTGAACGTCTGGTCCAGTATTTGAACGCGGTGCCCGAT GGCAGGATCCTTTCTGTTGAGTGAATGATGAAGGTTCTCGAAATCTGGATGACATGG CCAGGAAGGCGATGACCAATTTGGGAAGCAAACACTTCTGCACCTTGATTAGGGT GGAGTGGACGGAGTGGTTCGATCATGATAAAGTATCTCAGACTAAAGGTGGGGAGAAA ATTTAGACCTCTGGAAGCTCAGGAGGAAAAATATGCAATCGTCCCATTGATATAC AGCAGGCCACTACAATGGATGGAGTTAACTCAGCACCAGGAGGTGTCTACAAAAAGG CCAGGATTATAGGTTTGTCTGCTACGACCGGGCAGAGCCTGCCGGAGCTACCGGTGTA CGGTTCTCTGTGGGAAGCCTGTGAGGCCCAAACCTCACAGTCACCATTGACACCAATG		

	<p>TGAACAGCACCATTCTGAACTTGGAGGATAATGTACAGTCATGGAACTTGGAGATAC  CCTGGTCATTGCCAGTACTGATTACTCCATGTACCAGGCAGAAGTTCAGGTGCTT  CCCTGCAGATCCTGCGCCCCAACCCAGGTCAAAGTGGCAGGGAAACCAATGTACCTGC  ACATCGGGGAGGAGATAGACGGCGTGGACATGCGGGCGGAGGTTGGGCTTCTGAGCCG  GAACATCATAGTGATGGGGGAGATGGAGGACAAATGCTACCCCTACAGAAACCACATC  TGCAATTTCTTTGACTTCGATACCTTTGGGGGCCACATCAAGTTTGTCTGAGGATTTA  AGGCAGCACACTTGGAGGGCACGGAGCTGAAGCATATGGGACAGCAGCTGGTGGGTCA  GTACCCGATTCACTTCCACCTGGCCGGTGATGTAGACGAAAGGGGAGGTTATGACCCA  CCCACATACATCAGGGACCTCTCCATCCATCATACTTCTCTCGTGGCTCACAGTCC  ATGGCTCCAATGGCTTGTGTATCAAGGACGTTGTGGGCTATAACTCTTTGGGCCACTG  CTTCTTCACGGAAGATGGGCCGAGGAACGCAACACTTTTGACCACTGTCTTGGCCTC  CTTGTCAAGTCTGGAACCTCTCTCCCTCGGACCGTGACAGCAAGATGTGCAAGATGA  TCACAGAGGACTCCTACCCAGGGTACATCCCCAAGCCAGGCAAGACTGCAATGCTGT  GTCCACCTTCTGGATGGCCAATCCCAACAACAACCTCATCACTGTGCCGTGCAGGA  TCTGAGGAACTGGATTTTGGTTTATTTTACCACGTACCAACGGGCCCCCTCCGTGG  GAATGTACTCCCGAGGTATTTCAGAGCACATTCCTACTGGGAAAATTCTATAACACCG  AGCACATTCCAACCTACCGGGCTGGCATGATCATAGACAACGGAGTCAAAACCACCGAG  GCCTCTGCCAAGGACAAGCGGCCGTTCTCTCAATCATCTGTCCAGGATACAGCCCTC  ACCAGGACGCCGACCCGCTGAAGCCCCGGGAGCGGCCATCATCAGACACTTCATTGC  CTACAAGAACCAGGACCACGGGGCCTGGCTGCGCGCGGGGATGTGTGGCTGGACAGC  TGCCCGTTTGTGACAATGGCATTGGCCTGACCTGGCCAGTGGTGGAACTTCCCGT  ATGACGACGGCTCCAAGCAAGAGATAAAGAACAGCTTGTTTGTTGGCGAGAGTGGCAA  CGTGGGGACGGAATGATGGACAATAGGATCTGGGGCCCTGGCGGCTGGACCATAGC  GGAAGGACCCTCCCTATAGGCCAGAATTTTCAATTAGAGGAATCAGTTATATGATG  GCCCCATCAACATCCAAAACCTGCCTTTCCGAAAGTTTGTGGCCCTGGAGGGCCGCA  CACCAGCGCCCTGGCCTTCCGCTGAATAATGCCTGGCAGAGCTGCCCCATAACAAC  GTGACCGGCATTGCCTTTGAGGACGTTCCGATTACTTCCAGAGTGTTCTTCCGAGAGC  CTGGGCCCTGGTTCAACCAGCTGGACATGGATGGGGATAAGACATCTGTGTTCCATGA  CGTCGACGGCTCCGTGTCCGAGTACCCTGGCTCCTACCTCAGCAAGAAATGACAACCTGG  CTGGTCCGGCACCCAGACTGCATCAATGTTCCCGACTGGAGAGGGGCCATTGTCAGTG  GGTGCTATGCACAGATGTACATTCAAGCCTACAAGACCAGTAACCTGCGAATGAAGAT  CATCAAGAATGACTTCCCCAGCCACCCTCTTTACCTGGAGGGGGCGCTCACCAGGAGC  ACCCATTACCAGCAATACCAACCGGTTGTCAACCTGCAGAAAGGGCTACACCATCCACT  GGGACCAGACGGCCCCCGCGAAGTCCGCTCATCACTTCAACAAGGGCGA  CTGGATCCGAGTGGGGCTCTGTACCCGCGAGGCACCACATTCTCCATCCTCTCGGAT  GTTCAACAATCGCTGTGAAGCAAACGTCCAAGACGGGCGTCTTCTGTAGGACCTTGC  AGATGGACAAAGTGGAGCAGAGCTACCTGGCAGGAGCCACTACTACTGGGACGAGGA  CTCAGGGCTGTTGTTCTGAAGCTGAAAGCTCAGAACGAGAGAGAGAAGTTTGCTTTC  TGCTCCATGAAAGGCTGTGAGAGGATAAAGATTAAAGCTCTGATTCCAAGAACGCAG  GCGTCAGTGACTGCACGCCACAGCTTACCCCAAGTTTACCAGAGGGGCTGTCTGTAGA  CGTGCCGATGCCCAAGAAGCTCTTTGGTTCTCAGCTGAAAACAAAGGACCATTCTTG  GAGGTGAAGATGGAGAGTTCCAAGCAGCACTTCTTCCACCTCTGGAACGACTTCGCTT  ACATTGAAGTGGATGGGAAGAAGTACCCAGTTCGGAGGATGGCATCCAGGTGGTGGT  GATTGACGGGAACCAAGGGCGCGTGGTGAGCCACACGAGCTTCAGGAACCTCATTCTG  CAAGGCATACCATGGCAGCTTTTCAACTATGTGGCGACCATCCCTGACAATCCATAG  TGCTTATGGCATCAAAGGGAAGATACGTCTCCAGAGGCCCATGGACCAGAGTGCTGGA  AAAGCTTGGGGCAGACAGGGGTCTCAAGTTGAAAGAGCAAATGGCATTCTGTGGCTTC  AAAGGCAGCTTCCGGCCCCATCTGGGTGACACTGGACACTGAGGATCACAAGCCAAAA  TCTTCCAAGTTGTGCCATCCCTGTGGTGAAGAAGAAGAAGTTGTGAGGACAGCTGCC  GCCCCGTGCCACCTCGTGGTAGACTATGACGGTGAC</p>		
	ORF Start: TGC at 2	ORF Stop: TGA at 3815	
	SEQ ID NO: 190	1271 aa	MW at 143122.4kD
NOV42c, CG59889-04 Protein Sequence	<p>CPDQSPQLPWNPGHDQDHHVHIGQKTLTLLTSSATVYSIHISEGGKLVIKDHDEPIV  LRTRHILIDNGGELHAGSALCPFQGNFTIILYGRADEGIQDPYIYGLKYGVGKGGAL  ELHGQKKLSWTFNLKTLHPGMAEGGYFFERSWGHGVIHVHIDPKSGTVIHSDFDT  YRSKKESERLVQYLVNAPDGRILSVAVNDEGSRNLDDMARKAMTKLGSKHLHLGFRV  EWTEWFDHDKVSQTKGGEKISDLWKAHPGKICNRPIDIQATTMDGVNLTSEVYKKG  QDYRFACYDRGRACRSYRVFLCGKPVRLKLTVTIDTNVNSTILNLEDNVQSWKPGDT  LVIASDYSMYQAEFQVLPSCRSCAPNQVKVAGKPMYLIHIGEEIDGVMRAEVGLLSR</p>		

NIIVMGEMEDKCYPYRNHICNFFDFDTFGGHIKFALGFKAHLEGTELKHMGGQLVGO YPIHFHLAGDVDERGGYDPPTYIRDLSTHHTFSRCVTVHGSNGLLIKDVVGYNLSLGH FFTEDGPEERNTFDHCLGLLVKSGTLLPSDRDSKMCKMITEDSYPGYIPKPRQDCNAV STFWMANPNNLINCAAAGSEETGFWFIFHHVPTGPSVGMSPGYSEHILPGKFYNRR AHSNYRAGMIIDNGVKTTEASAKDKRPFLSIISARYSPHQDADPLKPREPAIIRHFIA YKNQDHGAWLRGGDVWLDSCRFDNGIGLTLASGGTFPYDDGSKQEIKNLSFVGESGN VGTEMDNRIWGPGLDHSGRTLPIGQNFPIRGIQLYDGPINIQNCTFRKFVALEGRH TSALAFRLNNAWQSCPHNNVTGIAFEDVPITSRVFFGEPGPWFNQDMDGDKTSVFHD VDGSVSEYPGSYLTKNLWLVHRPDCINVPDWRGAICSGCYAQMYIQAYKTSNLRMKI IKNDFPSHPLYLEGALTRSTHYQQYQPVVTLQKGYTIHWDQTAPAEALAIWLINFNKGD WIRVGLCYPRGTTFSILSDVHNRLKQTSKTGVFVRTLQMDKVEQSPGRSHYYWDED SGLLFLKLKAQNEREKFAFCSMKGCERIKIKALIPKNAGVSDCTATAYPKFTERAVVD VPMFKLFGSQLKTKDHFLEVMESSKQHFFHLWNDFAIYIEVDGKKYPSSSEDGIQVVV IDGNQGRVVSHTSFRNSILQGIQWQLFNYVATIPDNSIVLMASKGRYVSRGPWTRVLE KLGADRGLKLKEQMAFVGFGKSGFRPIWVTLDTEHAKAIFQVVPFVVKKKKL
---

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 42B.

Table 42B. Comparison of NOV42a against NOV42b through NOV42c.		
Protein Sequence	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV42b	103..1366	1257/1320 (95%)
	31..1349	1258/1320 (95%)
NOV42c	108..1366	1259/1259 (100%)
	1..1259	1259/1259 (100%)

Further analysis of the NOV42a protein yielded the following properties shown in Table 42C.

Table 42C. Protein Sequence Properties NOV42a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.6499 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.3000 probability located in nucleus
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 42D.

Table 42D. Geneseq Results for NOV42a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY25793				e-169

	encoded from gene 12 - Homo sapiens, 396 aa. [WO9938881-A1, 05-AUG-1999]	10..379	301/371 (80%)	
AAB67331	Human neuron progenitor cell clone #3 protein - Homo sapiens, 745 aa. [WO200107607-A2, 01-FEB-2001]	664..1357 1..701	311/711 (43%) 439/711 (61%)	e-169
AAG73990	Human colon cancer antigen protein SEQ ID NO:4754 - Homo sapiens, 194 aa. [WO200122920-A2, 05-APR-2001]	807..992 1..186	183/186 (98%) 184/186 (98%)	e-110
AAV25722	Human secreted protein encoded from gene 12 - Homo sapiens, 129 aa. [WO9938881-A1, 05-AUG-1999]	103..192 31..120	82/90 (91%) 82/90 (91%)	5e-43
AAV25801	Human secreted protein fragment encoded from gene 12 - Homo sapiens, 45 aa. [WO9938881-A1, 05-AUG-1999]	421..465 1..45	45/45 (100%) 45/45 (100%)	2e-18

In a BLAST search of public sequence databases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42E.

Table 42E. Public BLASTP Results for NOV42a				
Protein Accession Number	Protein/Organism/Length	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9ULM1	KIAA1199 PROTEIN - Homo sapiens (Human), 1013 aa (fragment).	365..1378 1..1013	1013/1014 (99%) 1013/1014 (99%)	0.0
AAH20256	HYPOTHETICAL 110.4 KDA PROTEIN - Homo sapiens (Human), 992 aa.	103..996 31..979	886/950 (93%) 887/950 (93%)	0.0
Q9NPN9	KIAA1199 HYPOTHETICAL PROTEIN - Homo sapiens (Human), 804 aa (fragment).	582..1378 8..804	797/797 (100%) 797/797 (100%)	0.0
Q9UHN6	TRANSMEMBRANE PROTEIN 2 - Homo sapiens (Human), 1383 aa.	1..1357 1..1339	622/1392 (44%) 843/1392 (59%)	0.0
Q9P2D5	KIAA1412 PROTEIN - Homo sapiens (Human), 1274 aa (fragment).	108..1357 4..1230	575/1275 (45%) 781/1275 (61%)	0.0

PFam analysis predicts that the NOV42a protein contains the domains shown in the Table 42F.



Table 42F. Domain Analysis of NOV42a			
Pfam Domain	NOV42a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found			

**EXAMPLE 43.**

The NOV43 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis			
	SEQ ID NO: 191	641 bp	
NOV43a, CG59512-02 DNA Sequence	AATCATGCAGGTCTCCACTGCTGCCCTTGCTGTCCCCCTCTGCACCATGGCTCTCTGCAACCAGT TCTCTGCATCACTTGCTGCTGACACGCCGACCGCTGCTGCTTCAGCTACACCTCCCGGCAGATT CCACAGAATTTATAGCTGACTACTTTGAGACGAGCAGCCAGTGCTCCAAGCCCGGTGTCATCTT CCTAACCAAGAGAGGCCCGGCAGGTCTGTGCTGACCCAGTGAGGAGTGGGTCCAGAAATACGTCA GTGACCTGGAGCTGAGTGCCTGAG		
	ORF Start: ATG at 5	ORF Stop: TGA at 284	
	SEQ ID NO: 192	92 aa	MW at 10039.3kD
NOV43a, CG59512-02 Protein Sequence	MQVSTAALAVPLCTMALCNQFSASLAADTPTACCFSTSRQIPQNFIADYFETSSQCSK PGVIFLTRGRQVCADPSEEWVQKYVSDLELSA		
	SEQ ID NO: 193	638 bp	
NOV43b, CG59512-01 DNA Sequence	AATCATGCAGGTCTCCACTGCTGTCTTGCTGTCTCTCTGACCATGGCTCTCTGC AACCAGTTCTCTGCATCACTTGCTGCTGACACGCCGACCGCTGCTGCTTCAGCTACA CCTCCCGGCAGATTCCACAGAATTTATAGCTGACTACTTTGAGACGAGCAGCCAGTG CTCCAAGCCCGGTGTCATCTTCTAACCAAGCGAAGCCCGGCAGGTCTGTGCTGACCCC AGTGAGGAGTGGGTCCAGAAATATGTCAGCGACCTGGAGCTGAGTGCCTGAGGGGTCC AGAAGCTTCGAGGCCAGCGACCTCGGTGGGCCAGTGAGGAGGAGCAGGAGCCTGAG CCTTGGGAACATGCGTGTGACCTCCACAGCTACCTCTTCTATGGACTGGTTGTTGCCA AACAGCCACACTGTGGGACTCTTCTTAACCAAGCGAAGCCCGGCAGGTCTGTGCTGACC CCAGTGAGGAGTGGGTCCAGAAATATGTCAGCGACCTGGAGCTGAGTGCCTGAGGGGT CCAGAAGCTTCGAGGCCAGCGACCTCGGTGGGCCAGTGAGGAGGAGCAGGAGCCTG AGCCTTGGGAACATGCGTGTGACCTCCACAGCTACCTCTTCTATGGACTGGTTGTTGC		
	ORF Start: ATG at 5	ORF Stop: TGA at 281	
	SEQ ID NO: 194	92 aa	MW at 10113.4kD
NOV43b, CG59512-01 Protein Sequence	MQVSTAVLAVLLCTMALCNQFSASLAADTPTACCFSTSRQIPQNFIADYFETSSQCS KPGVIFLTRSRQVCADPSEEWVQKYVSDLELSA		

Sequence comparison of the above protein sequences yields the following sequence  
5 relationships shown in Table 43B.

Table 43B. Comparison of NOV43a against NOV43b and NOV43c.		
Protein Sequence	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV43b	1..92	89/92 (96%)
	1..92	89/92 (96%)

Further analysis of the NOV43a protein yielded the following properties shown in Table 43C.

Table 43C. Protein Sequence Properties NOV43a	
PSort analysis:	0.6997 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Likely cleavage site between residues 28 and 29

- A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several
- 5 homologous proteins shown in Table 43D.

Table 43D. Geneseq Results for NOV43a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB11876	Human G0S19-2 peptide precursor homologue, SEQ ID NO:2246 - Homo sapiens, 124 aa. [WO200157188-A2, 09-AUG-2001]	1..93 32..124	90/93 (96%) 90/93 (96%)	5e-47
AAU09185	Human PRO10008 polypeptide - Homo sapiens, 93 aa. [WO200166740-A2, 13-SEP-2001]	1..93 1..93	90/93 (96%) 90/93 (96%)	5e-47
AAY96281	Human chemokine MIP-1alpha - Homo sapiens, 93 aa. [WO200028035-A1, 18-MAY-2000]	1..93 1..93	90/93 (96%) 90/93 (96%)	5e-47
AAB15807	Human chemokine C10 SEQ ID NO: 49 - Homo sapiens, 93 aa. [WO200042071-A2, 20-JUL-2000]	1..93 1..93	90/93 (96%) 90/93 (96%)	5e-47
AAW82721	Human MI10 protein - Homo sapiens, 93 aa. [WO9854326-A1, 03-DEC-1998]	1..93 1..93	90/93 (96%) 90/93 (96%)	5e-47

In a BLAST search of public sequence databases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43E.

Table 43E. Public BLASTP Results for NOV43a
---

Protein Accession Number	Protein/Organism/Length	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P16619	Small inducible cytokine A3 like 1 precursor (Tonsillar lymphocyte LD78 beta protein) (G0/G1 switch regulatory protein 19-2) (G0S19-2 protein) (PAT 464.2) - Homo sapiens (Human), 93 aa.	1..93 1..93	90/93 (96%) 90/93 (96%)	2e-46
P10147	Small inducible cytokine A3 precursor (Macrophage inflammatory protein 1-alpha) (MIP-1-alpha) (Tonsillar lymphocyte LD78 alpha protein) (G0/G1 switch regulatory protein 19-1) (G0S19-1 protein) (SIS-beta) (PAT 464.1) - Homo sapiens (Human), 92 aa.	1..93 1..92	91/93 (97%) 91/93 (97%)	7e-46
Q96168	SIMILAR TO SMALL INDUCIBLE CYTOKINE A3 (HOMOLOGOUS TO MOUSE MIP-1A) - Homo sapiens (Human), 93 aa.	1..93 1..93	89/93 (95%) 89/93 (95%)	1e-45
Q14745	LD78 ALPHA BETA PRECURSOR - Homo sapiens (Human), 80 aa (fragment).	7..87 1..80	76/81 (93%) 77/81 (94%)	5e-38
P50229	Small inducible cytokine A3 precursor (Macrophage inflammatory protein 1-alpha) (MIP-1-alpha) - Rattus norvegicus (Rat), 92 aa.	1..93 1..92	71/93 (76%) 82/93 (87%)	1e-35

Pfam analysis predicts that the NOV43a protein contains the domains shown in the Table 43F.

Table 43F. Domain Analysis of NOV43a			
Pfam Domain	NOV43a Match Region	Identities/ Similarities for the Matched Region	Expect Value
IL8: domain 1 of 1	24..89	31/70 (44%) 62/70 (89%)	6e-34

**EXAMPLE 44.**

The NOV44 clone was analyzed, and the nucleotide and predicted polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis		
	SEQ ID NO: 195	1737 bp

NOV44a, CG56801-02 DNA Sequence	CATGCTTGGGGTCTGGTCCTTGGCGCGCTGGCCCTGGCCGGCCTGGGGCTCCCCGCA CCCGCAGAGCCCGAGCCGGGTGGCAGCCAGTGCCTCGAGCAGCACTGCTTCGCGCTCT ACCCGGGCCCCCGACCTTCTCAATGCCAGTCAGATCTGCGACGGACTGCGGGGCCA CCTAATGACAGTGCCTCTCGGTGGCTGCCGATGTCATTTCTTGTACTGAACGGC GACGGCGGCGTTGGCCGCCGGCGCCTCTGGATCGGCCTGCAGCTGCCACCCGGCTGCG GCGACCCCAAGCGCTCGGGCCCCCTGCGCGGCTTCCAGTGGGTACGGGAGACAACAA CACCAGCTATAGCAGGTGGGCACGGCTCGACCTCAATGGGGTCCCCCTCTGCGGCCG TTGTGCGTCGCTGTCTCCGCTGCTGAGGCCACTGTGCCCAGCGAGCCGATCTGGGAGG AGCAGCAGTGCGAAGTGAAGGCCGATGGCTTCCTCTGCGAGTTCCACTTCCAGCCAC CTGCAGGCCACTGGCTGTGGAGCCCGCGCCGGCTGCCCGCTCTCGATCACCTAC GGCACCCCGTTCCGCGCCCGCGGAGCGGGCTTCCAGGCGCTGCCGGTGGGCAGCTCCG CCGCGGTGGCTCCCCCTCGGCTTACAGCTAATGTGCACCGCGCCGCCCGGAGCGGTCCA GGGGCACTGGGCCAGGGAGCGCCGGCGCTTGGGACTGCAGCTGGAGAACGGCGGC TGCGAGCACACGTGCAATGCGATCCCTGGGGCTCCCCGCTGCCAGTGCCAGCCGGCG CCGCCCTGCAGGCAGACGGGCGCTCCTGCACCGCATCCGCGACGAGTCTTGCACGA CCTCTGCGAGCACTTCTGCGTTCCCAACCCCGACAGCCGGGCTCCTACTCGTGCATG TGCGAGACCGGCTACCGGCTGGCGGCCGACCAACACCGGTGCGAGGACGTGGATGACT GCATACTGGAGCCAGTCCGTGTCCGAGCGCTGTGTCAACACACAGGGTGGCTTCCA GTGCCACTGCTACCCCTAACTACGACCTGGTGGACGGCGAGTGTGTGGAGCCCGTGGAC CCGTGCTTCAGAGCCAACCTGCGAGTACCAGTGCAGCCCTGAACCAAAGTACTACC TCTGCGTCTGCGCCGAGGGCTTCGCGCCCATTCCCACGAGCCGACAGGTGCCAGAT GTTTTGCAACAGACTGCCTGTCCAGCCGACTGCGATCCCAACACCCAGGCTAGCTGT GAGTGCCCTGAAGGCTACATCCTGGACGACGGTTTCATCTGCACGGACATCGCAGAT GCGAAAACGGCGGCTTCTGCTCCGGGGTGTGCCACAACCTCCCGGTGCTTCCAGTG CATCTGCGGGCCGACTCGGCCCTTGCCCGCCACATTGGCACCGACTGTGACTCCGGC AAGGTGGACGGTGGCGACAGCGGCTCTGGCGAGCCCCCGCCAGCCGACGCGCGGCT CCACCTTGACTCCTCCGGCGTGGGGCTCGTGCATTCGGGCTTGCTCATAGGCATCTC CATCGCGAGCCTGTGCTGGTGGTGGCGCTTTTGGCGCTCCTTGCCACCTGCGCAAG AAGCAGGGCGCCCGCAGGGCCAAGATGGAGTACAAGTGCAGCGCCCTTCCAAGGAGG TAGTGCTGCAGCACGTGCGGACCGAGCGGACGCCGAGAGACTCTAGCGGCCTCC		
	ORF Start: ATG at 2	ORF Stop: TAG at 1727	
	SEQ ID NO: 196	575 aa	MW at 60266.5kD
NOV44a, CG56801-02 Protein Sequence	MLGVLVLGALALAGLGLPAPAEPPQPGGSQCVHDFALYPGPATFLNASQICDGLRGH LMTVRSSVAADVLSLLNGDGGVGRRLWIGLQLPPGCGDPKRLGFLRGFQVWTDNN TSYSRWARLDLNGAPLCGPLCAVSAAEATVPSEPIWEEQQCEVKADGFLCEFHFPAT CRPLAVEPGAAAAVSIITYGTFPAARGAGFQALPVGSSAAVAPLGLQLMCTAPPGAVQ GHWAREAPGAWDCSVENGCEHTCNAIPGAPRCQCPAGAALQADGRSCTASATQSCND LCEHFCVPNPDPGYSYSCMCETGYRLAADQHRCEVDVDDCILEPSPCPQRCVNTQGGFE CHCYPNYDLVDGECVEPVPDPCFRANCEYQCQLNQTSYLCVCAEGFAPIPHEPHRCQM FCNQTACPADCDPNTQASCECEGYILDGFICTDIDECENGFCSGVCHNLPGTFEC ICGPDALARHIGTDGDSGKVDGGDSGSGEPSPSTPGSTLTPPAVGLVHSGLLIGIS IASLCLVVALLALLCHLRKKQGAARAKMEYKCAAPSKEVVLQHVRTERTPQRL		

Further analysis of the NOV44a protein yielded the following properties shown in Table 44B.

Table 44B. Protein Sequence Properties NOV44a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Likely cleavage site between residues 24 and 25

A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publications, yielded several homologous proteins shown in Table 44C.

<b>Table 44C. Geneseq Results for NOV44a</b>				
<b>Geneseq Identifier</b>	<b>Protein/Organism/Length [Patent #, Date]</b>	<b>NOV44a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Region</b>	<b>Expect Value</b>
AAR43031	Human thrombomodulin - Homo sapiens, 575 aa. [WO9322447-A, 11-NOV-1993]	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0
AAR41806	Thrombomodulin - Homo sapiens, 575 aa. [JP05213998-A, 24-AUG-1993]	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0
AAR11534	Human thrombomodulin type II polypeptide, 575 aa. [WO9104276-A, 04-APR-1991]	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0
AAP82070	Human thrombomodulin encoded by plasmid p2.1 - synthetic, 575 aa. [WO8809811-A, 15-DEC-1988]	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0
AAR31572	Human thrombomodulin - Synthetic, 575 aa. [WO9301282-A, 21-JAN-1993]	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0

- In a BLAST search of public sequence databases, the NOV44a protein was found to
- 5 have homology to the proteins shown in the BLASTP data in Table 44D.

<b>Table 44D. Public BLASTP Results for NOV44a</b>				
<b>Protein Accession Number</b>	<b>Protein/Organism/Length</b>	<b>NOV44a Residues/ Match Residues</b>	<b>Identities/ Similarities for the Matched Portion</b>	<b>Expect Value</b>
P07204	Thrombomodulin precursor (Fetomodulin) (TM) (CD141 antigen) - Homo sapiens (Human), 575 aa.	1..575 1..575	572/575 (99%) 572/575 (99%)	0.0
THHUB	thrombomodulin precursor [validated] - human, 575 aa.	1..575 1..575	571/575 (99%) 571/575 (99%)	0.0
Q9UC32	THROMBOMODULIN - Homo sapiens (Human), 468 aa.	19..486 1..468	465/468 (99%) 465/468 (99%)	0.0
P15306				0.0

	(Fetomodulin) (TM) - Mus musculus (Mouse), 577 aa.	1..576	443/579 (76%)	
O35370	THROMBOMODULIN - Rattus norvegicus (Rat), 577 aa.	1..574 1..576	378/578 (65%) 435/578 (74%)	0.0

PFam analysis predicts that the NOV44a protein contains the domains shown in the Table 44E.

Table 44E. Domain Analysis of NOV44a			
Pfam Domain	NOV44a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lectin_c: domain 1 of 1	41..169	27/138 (20%) 86/138 (62%)	0.0032
EGF: domain 1 of 6	245..280	14/47 (30%) 28/47 (60%)	1.1e-05
EGF: domain 2 of 6	288..323	14/47 (30%) 26/47 (55%)	0.022
metalthio: domain 1 of 1	261..325	15/73 (21%) 39/73 (53%)	9
EGF: domain 3 of 6	329..362	13/47 (28%) 24/47 (51%)	1.6
EGF: domain 4 of 6	369..404	9/47 (19%) 23/47 (49%)	1.5
EB: domain 1 of 1	351..404	15/61 (25%) 36/61 (59%)	4.8
EGF: domain 5 of 6	408..439	11/47 (23%) 19/47 (40%)	9.4
EGF: domain 6 of 6	445..480	12/47 (26%) 25/47 (53%)	0.7

## 5 EXAMPLE 45: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological

states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

**2. SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

**3. PathCalling™ Technology:**

The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected



by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

5     **4. RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these  
10    procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5.     **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by  
15    starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in  
20    silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain -  
25    thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector  
30    from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In

addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

15 **Example 46: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences**

- Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the

relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs  
5 used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included  
10 because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program  
15 SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST  
20 locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

25 Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

#### **NOV2a SNP data:**

30 NOV2a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:9 and 10, respectively. The nucleotide sequence of the NOV2a variant differs as shown in Table 46A.

<b>Table 46A SNP data for NOV2a</b>
-------------------------------------

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377289	384	T	C	125	His	His
13377288	405	C	T	132	Val	Val
13377287	672	C	T	221	Val	Val

**NOV6a SNP data:**

NOV6a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:23 and 24, respectively. The nucleotide sequence of the NOV6a variant differs as shown in Table 46B.

Table 46B SNP data for NOV6a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377290	1592	G	T	519	Ala	Ala
13377291	2089	T	C	685	Ile	Thr

**NOV7a SNP data:**

NOV7a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:25 and 26, respectively. The nucleotide sequence of the NOV7a variant differs as shown in Table 46C.

Table 46C SNP data for NOV7a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374597	67	C	T	22	Pro	Leu
13374596	129	C	T	43	Gln	End
13374595	267	C	T	89	Pro	Ser

**NOV9a SNP data:**

NOV9a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:31 and 32, respectively. The nucleotide sequence of the NOV9a variant differs as shown in Table 46D.

Table 46D SNP data for NOV9a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374168	81	C	G	27	Pro	Pro
13374236	160	C	A	54	Arg	Arg
13374237	192	G	A	64	Gly	Gly
13375849	355	A	G	119	Asn	Asp

**NOV11a SNP data:**

- 5 NOV11a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:35 and 36, respectively. The nucleotide sequence of the NOV11a variant differs as shown in Table 46E.

Table 46E SNP data for NOV11a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377303	124	T	C	14	Phe	Leu
13377301	858	C	T	258	Tyr	Tyr
13377300	868	A	G	262	Ser	Gly
13377299	951	G	A	289	Trp	End

10 **NOV14a SNP data:**

NOV14a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:57 and 58, respectively. The nucleotide sequence of the NOV14a variant differs as shown in Table 46F.

15

Table 46F SNP data for NOV14a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374670	92	C	A	17	Ala	Glu
13374669	146	A	G	35	Glu	Gly
13374668	247	T	C	69	Phe	Leu

13374667	266	C	T	75	Ala	Val
----------	-----	---	---	----	-----	-----

**NOV15a SNP data:**

NOV15a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:59 and 60, respectively. The nucleotide sequence of the NOV15a variant differs as shown in Table 46G.

Table 46G SNP data for NOV15a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377304	21	A	G	2	Arg	Gly
13374822	256	G	T	80	Trp	Leu

**NOV16a SNP data:**

NOV16a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:65 and 66, respectively. The nucleotide sequence of the NOV16a variant differs as shown in Table 46H.

Table 46H SNP data for NOV16a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377305	301	C	T	92	Ala	Ala
13374717	942	G	A	306	Arg	Gln
13377306	1183	T	C	386	Gly	Gly
13377307	1503	C	T	493	Ser	Phe

**NOV18a SNP data:**

NOV18a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:73 and 74, respectively. The nucleotide sequence of the NOV18a variant differs as shown in Table 46H.

Table 46H SNP data for NOV18a
-------------------------------

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377309	951	C	T	306	Arg	Trp

**NOV21a SNP data:**

NOV21a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:85 and 86, respectively. The nucleotide sequence of the NOV21a variant differs as shown in Table 46I.

Table 46I SNP data for NOV21a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13374712	373	A	G	84	Ile	Val

**NOV25a SNP data:**

10

NOV25a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:121 and 122, respectively. The nucleotide sequence of the NOV25a variant differs as shown in Table 46J.

Table 46J SNP data for NOV25a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377310	361	C	T	117	Ser	Ser

15

**NOV27a SNP data:**

NOV27a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:127 and 128, respectively. The nucleotide sequence of the NOV27a variant differs as shown in Table 46K.

Table 46K SNP data for NOV27a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified

13377311	159	T	C	45	Trp	Arg
13377314	671	C	T	215	Gly	Gly
13377312	739	A	G	238	Tyr	Cys
13377313	774	A	G	250	Thr	Ala

**NOV28a SNP data:**

NOV28a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:129 and 130, respectively. The nucleotide sequence of the NOV28a variant differs as shown in Table 46K.

Table 46K SNP data for NOV28a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377318	145	T	G	41	Pro	Pro
13377317	162	A	G	47	His	Arg
13375785	351	A	G	110	Glu	Gly
13375450	411	T	C	130	Leu	Pro
13377316	577	C	T	185	Ala	Ala
13377315	968	G	A	316	Gly	Arg
13375452	990	A	G	323	Glu	Gly

**NOV31a SNP data:**

10

NOV31a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:141 and 142, respectively. The nucleotide sequence of the NOV31a variant differs as shown in Table 46L.

Table 46L SNP data for NOV31a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377319	1221	A	G	371	Thr	Ala

15

**NOV34a SNP data:**



NOV34a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:147 and 148, respectively. The nucleotide sequence of the NOV34a variant differs as shown in Table 46M.

Table 46M SNP data for NOV34a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377321	240	T	C	80	Ser	Ser
13377320	492	T	C	164	Asp	Asp

5

#### NOV40a SNP data:

NOV40a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:179 and 180, respectively. The nucleotide sequence of the NOV40a variant differs as shown in Table 46N.

10

Table 46N SNP data for NOV40a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13376614	1732	G	A	561	Gly	Asp
13376613	3266	C	T	1072	Phe	Phe
13376612	4183	A	G	1378	Asp	Gly
13376611	4604	C	T	1518	Gly	Gly
13376610	4625	C	T	1525	Asp	Asp
13376609	5491	T	C	1814	Leu	Pro
13376596	5589	C	T	1847	Gln	End
13376597	5637	T	C	1863	Ser	Pro
13376608	5765	T	C	1905	Asp	Asp
13376607	6469	A	G	0		

#### NOV42a SNP data:

NOV42a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:185 and 186, respectively. The nucleotide sequence of the NOV42a variant differs as shown in Table 46O.

15

Table 46O SNP data for NOV42a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377323	2186	G	A	722	Gly	Asp
13377322	3820	G	T	1267	Val	Leu

**NOV44a SNP data:**

- 5 NOV44a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:195 and 196, respectively. The nucleotide sequence of the NOV44a variant differs as shown in Table 46P.

Table 46P SNP data for NOV44a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13375190	50	C	T	17	Leu	Phe
13374613	764	A	G	255	Thr	Ala
13374614	1413	C	T	471	Ala	Val
13375192	1419	C	T	473	Ala	Val

10 **Example 47. Quantitative expression analysis of clones in various cells and tissues**

- The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System.
- 15 Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on
- 20 metabolic diseases), AI\_comprehensive\_panel (containing normal tissue and samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from

normal and diseased brains) and CNS\_neurodegeneration\_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example,  $\beta$ -actin and GAPDH). Normalized RNA (5  $\mu$ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10  $\mu$ g of total RNA were performed in a volume of 20  $\mu$ l and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50  $\mu$ g of total RNA in a final volume of 100  $\mu$ l. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature ( $T_m$ ) range = 58°-60°C, primer optimal  $T_m$  = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe  $T_m$  must be 10°C greater than primer  $T_m$ , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthesgen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate

(Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using IX TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

#### **Panels 1, 1.1, 1.2, and 1.3D**

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

\* = established from metastasis,

met = metastasis,

5 s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

10 astro = astrocytoma, and

neuro = neuroblastoma.

#### **General\_screening\_panel\_v1.4**

The plates for Panel 1.4 include 2 control wells (genomic DNA control and chemistry  
 15 control) and 94 wells containing cDNA from various samples. The samples in Panel 1.4 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer.  
 20 Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panel 1.4 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal  
 25 skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.  
 Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

30

#### **Panels 2D and 2.2**

The plates for Panels 2D and 2.2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network

(CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

15

#### **Panel 3D**

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

25

#### **Panels 4D, 4R, and 4.1D**

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients

30

was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

5 Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

15 Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2µg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5µg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately  $2 \times 10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ( $5.5 \times 10^{-5}$ M) (Gibco), and 10mM Hepes (Gibco).

25 The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum

- (FCS) (Hyclone, Logan, UT), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 $\mu$ g/ml for 6 and 12-14 hours.
- 10 CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO
- 15 beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), and 10mM Hepes (Gibco) and plated at 10<sup>6</sup> cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight
- 20 with 0.5 $\mu$ g/ml anti-CD28 (Pharmingen) and 3 $\mu$ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco),
- 25 mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco),
- 30 mercaptoethanol 5.5x10<sup>-5</sup>M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10<sup>6</sup> cells/ml in DMEM 5% FCS (Hyclone), 100 $\mu$ M non essential amino acids



(Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5µg/ml or anti-CD40 (Pharmingen) at approximately 10µg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24, 48 and 72 hours.

5 To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10µg/ml anti-CD28 (Pharmingen) and 2µg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at  $10^5$ - $10^6$  cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$  M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1µg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1µg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate  
10 (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1µg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes  
15 were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1,  
25 KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at  $5 \times 10^5$  cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to  $5 \times 10^5$  cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1µg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol  $5.5 \times 10^{-5}$ M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were  
30

activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately  
5 10<sup>7</sup> cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at  
10 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300µl of RNase-free water and 35µl buffer (Promega) 5µl DTT, 7µl RNAsin and 8µl DNase were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with  
15 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80°C.

#### **AI\_comprehensive panel\_v1.0**

The plates for AI\_comprehensive panel\_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained  
20 from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue  
25 samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as  
30 total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three

female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebid and two were on phenobarbital.

- 5           Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1 anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that  
10       could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI\_comprehensive panel\_v1.0 panel, the following abbreviations are used:

- AI = Autoimmunity  
15       Syn = Synovial  
Normal = No apparent disease  
Rep22 /Rep20 = individual patients  
RA = Rheumatoid arthritis <  
Backus = From Backus Hospital  
20       OA = Osteoarthritis  
(SS) (BA) (MF) = Individual patients  
Adj = Adjacent tissue  
Match control = adjacent tissues  
-M = Male  
25       -F = Female  
COPD = Chronic obstructive pulmonary disease

#### **Panels 5D and 5I**

- The plates for Panel 5D and 5I include two control wells and a variety of cDNAs  
30       isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2: Diabetic Hispanic, overweight, not on insulin

Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)

Patient 10: Diabetic Hispanic, overweight, on insulin

Patient 11: Nondiabetic African American and overweight

Patient 12: Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the

University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

- 5 GO Adipose = Greater Omentum Adipose
- SK = Skeletal Muscle
- UT = Uterus
- PL = Placenta
- AD = Adipose Differentiated
- 10 AM = Adipose Midway Differentiated
- U = Undifferentiated Stem Cells

#### **Panel CNSD.01**

- The plates for Panel CNSD.01 include two control wells and 94 test samples
- 15 comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

- 20 Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal
- 25 cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were
- 30 examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

5 Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

#### **Panel CNS\_Neurodegeneration\_V1.0**

The plates for Panel CNS\_Neurodegeneration\_V1.0 include two control wells and 47  
 10 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from  
 the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and  
 Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are  
 removed from calvaria of donors between 4 and 24 hours after death, sectioned by  
 neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and  
 15 examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from  
 Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no  
 evidence of dementia prior to death. The eight normal control brains are divided into two  
 categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and  
 20 controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically  
 senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD  
 senile plaque load). Within each of these brains, the following regions are represented:  
 hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and  
 occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of  
 25 neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in  
 AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus;  
 the parietal cortex shows moderate neuronal death in the late stages of the disease; the  
 occipital cortex is spared in AD and therefore acts as a "control" region within AD patients.  
 Not all brain regions are represented in all cases.

30 In the labels employed to identify tissues in the CNS\_Neurodegeneration\_V1.0 panel,  
 the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology  
 upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; pateint not demented but showing sever AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

5

#### A. NOV1a, NOV1c, and NOV1d: NEUREXOPHILIN 1 PRECURSOR

Expression of gene NOV1a and variants NOV1c and NOV1d was assessed using the primer-probe set Ag3371, described in Table AA. Results of the RTQ-PCR runs are shown in Tables AB and AC. Please note that NOV1c and NOV1d represent full-length physical clones of the NOV1a gene, validating the prediction of the gene sequence.

10

Table AA. Probe Name Ag3371

Primers	Sequences	Length	Start Position
Forward	5'-acatatggacagaaagcagcaa-3' (SEQ ID NO:197)	22	114
Probe	TET-5'-ttgtctatcagccgaactcctgtcaca-3'-TAMRA (SEQ ID NO:198)	26	140
Reverse	5'-tatcattctctttgccacgaaa-3' (SEQ ID NO:199)	22	170

Table AB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3371, Run 210154070	Tissue Name	Rel. Exp.(%) Ag3371, Run 210154070
AD 1 Hippo	7.3	Control (Path) 3 Temporal Ctx	2.1
AD 2 Hippo	25.5	Control (Path) 4 Temporal Ctx	63.7
AD 3 Hippo	3.5	AD 1 Occipital Ctx	13.0
AD 4 Hippo	8.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	86.5	AD 3 Occipital Ctx	3.1
AD 6 Hippo	23.2	AD 4 Occipital Ctx	37.9
Control 2 Hippo	28.3	AD 5 Occipital Ctx	56.6
Control 4 Hippo	5.7	AD 6 Occipital Ctx	13.5
Control (Path) 3 Hippo	4.0	Control 1 Occipital Ctx	2.0
AD 1 Temporal Ctx	9.9	Control 2 Occipital Ctx	46.0
AD 2 Temporal Ctx	42.0	Control 3 Occipital Ctx	13.9
AD 3 Temporal Ctx	2.9	Control 4 Occipital Ctx	4.5
AD 4 Temporal Ctx	21.3	Control (Path) 1 Occipital Ctx	90.1
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.2
AD 5 Sup Temporal Ctx	39.0	Control (Path) 3 Occipital Ctx	1.0
AD 6 Inf Temporal Ctx	26.4	Control (Path) 4 Occipital Ctx	27.0
AD 6 Sup Temporal	31.2	Control 1 Parietal Ctx	6.0

Ctx			
Control 1 Temporal Ctx	7.0	Control 2 Parietal Ctx	26.1
Control 2 Temporal Ctx	64.6	Control 3 Parietal Ctx	15.5
Control 3 Temporal Ctx	19.5	Control (Path) 1 Parietal Ctx	82.4
Control 3 Temporal Ctx	6.3	Control (Path) 2 Parietal Ctx	54.3
Control (Path) 1 Temporal Ctx	69.3	Control (Path) 3 Parietal Ctx	5.8
Control (Path) 2 Temporal Ctx	38.4	Control (Path) 4 Parietal Ctx	46.0

Table AC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3371, Run 217043080	Tissue Name	Rel. Exp.(%) Ag3371, Run 217043080
Adipose	0.5	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.6
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.9	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.1	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.2	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.6
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.2
Ovarian ca. OVCAR-5	2.9	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.7	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.3	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.3
Breast ca. MCF-7	1.1	Heart Pool	0.0
Breast ca. MDA-MB- 231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.1
Breast ca. T47D	2.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	4.9
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.1
Lung	0.1	CNS cancer (glio/astro) U- 118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met)	0.0



		SK-N-AS	
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.2	CNS cancer (astro) SNB-75	0.3
Lung ca. NCI-H146	34.2	CNS cancer (glio) SNB-19	0.4
Lung ca. SHP-77	7.4	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	40.9
Lung ca. NCI-H526	0.2	Brain (cerebellum)	26.8
Lung ca. NCI-H23	2.4	Brain (fetal)	100.0
Lung ca. NCI-H460	6.8	Brain (Hippocampus) Pool	50.3
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	56.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	62.4
Liver	0.1	Brain (Thalamus) Pool	71.2
Fetal Liver	0.0	Brain (whole)	55.9
Liver ca. HepG2	0.1	Spinal Cord Pool	17.9
Kidney Pool	0.0	Adrenal Gland	49.3
Fetal Kidney	0.9	Pituitary gland Pool	2.6
Renal ca. 786-0	0.0	Salivary Gland	0.7
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3371 This panel confirms the expression of this gene at moderate levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased  
5 postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3371 Moderate expression of the NOV1a gene is seen in all regions of the brain represented on this panel (CT=29.2-31.7), with  
10 the highest level of expression in fetal brain. Thus, expression of this gene may be used to distinguish brain from the other samples on this panel. The NOV1a gene encodes a protein with homology to neurexophilins. Neurexophilins are members of a family of neuropeptide-like glycoproteins that bind to alpha-neurexins, receptor-like proteins expressed on the neuronal cell surface (Missler and Sudhof, J Neurosci 18(10):3630-8, 1998). Therefore, this  
15 gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4D Summary:** Ag3371 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**B. NOV2a: NEUROPHILIN**

Expression of gene NOV2a was assessed using the primer-probe set Ag3369, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB, BC and BD.

**Table BA.** Probe Name Ag3369

Primers	Sequences	Length	Start Position
Forward	5'-gtccacttccaacacaatgc-3' (SEQ ID NO:200)	20	403
Probe	TET-5'-agggaaacatctccatcagcctcgt-3'-TAMRA (SEQ ID NO:201)	25	431
Reverse	5'-ctgttcctcgtggaactctaca-3' (SEQ ID NO:202)	22	471

5

**Table BB.** CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3369, Run 210153743	Tissue Name	Rel. Exp.(%) Ag3369, Run 210153743
AD 1 Hippo	8.5	Control (Path) 3 Temporal Ctx	4.1
AD 2 Hippo	24.1	Control (Path) 4 Temporal Ctx	36.3
AD 3 Hippo	4.5	AD 1 Occipital Ctx	14.9
AD 4 Hippo	6.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	82.9	AD 3 Occipital Ctx	13.9
AD 6 Hippo	17.4	AD 4 Occipital Ctx	27.7
Control 2 Hippo	28.5	AD 5 Occipital Ctx	52.5
Control 4 Hippo	7.5	AD 6 Occipital Ctx	21.3
Control (Path) 3 Hippo	5.0	Control 1 Occipital Ctx	7.7
AD 1 Temporal Ctx	10.3	Control 2 Occipital Ctx	100.0
AD 2 Temporal Ctx	25.2	Control 3 Occipital Ctx	30.1
AD 3 Temporal Ctx	4.4	Control 4 Occipital Ctx	11.5
AD 4 Temporal Ctx	19.2	Control (Path) 1 Occipital Ctx	72.7
AD 5 Inf Temporal Ctx	61.6	Control (Path) 2 Occipital Ctx	22.7
AD 5 Sup Temporal Ctx	30.6	Control (Path) 3 Occipital Ctx	4.4
AD 6 Inf Temporal Ctx	22.1	Control (Path) 4 Occipital Ctx	31.0
AD 6 Sup Temporal Ctx	24.1	Control 1 Parietal Ctx	11.9
Control 1 Temporal Ctx	8.4	Control 2 Parietal Ctx	22.4
Control 2 Temporal Ctx	36.9	Control 3 Parietal Ctx	22.7
Control 3 Temporal Ctx	15.8	Control (Path) 1 Parietal Ctx	70.7
Control 3 Temporal Ctx	9.2	Control (Path) 2 Parietal Ctx	37.6
Control (Path) 1 Temporal Ctx	48.6	Control (Path) 3 Parietal Ctx	6.9
Control (Path) 2 Temporal Ctx	32.8	Control (Path) 4 Parietal Ctx	48.3

Table BC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3369, Run 217042734	Tissue Name	Rel. Exp.(%) Ag3369, Run 217042734
Adipose	2.6	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	4.1	Bladder	0.3
Melanoma* Hs688(B).T	9.1	Gastric ca. (liver met.) NCI-N87	0.2
Melanoma* M14	0.1	Gastric ca. KATO III	0.0
Melanoma* LOXIMV1	0.0	Colon ca. SW-948	0.1
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	3.5
Squamous cell carcinoma SCC-4	0.1	Colon ca.* (SW480 met) SW620	1.3
Testis Pool	2.1	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.1	Colon ca. HCT-116	0.3
Prostate Pool	2.7	Colon ca. CaCo-2	1.8
Placenta	1.7	Colon cancer tissue	1.4
Uterus Pool	5.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.7	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.4	Colon ca. SW-48	0.9
Ovarian ca. OVCAR-4	0.9	Colon Pool	19.1
Ovarian ca. OVCAR-5	2.0	Small Intestine Pool	10.7
Ovarian ca. IGROV-1	0.0	Stomach Pool	4.1
Ovarian ca. OVCAR-8	0.2	Bone Marrow Pool	5.8
Ovary	2.3	Fetal Heart	0.7
Breast ca. MCF-7	1.0	Heart Pool	7.2
Breast ca. MDA-MB- 231	0.1	Lymph Node Pool	13.5
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	0.7
Breast ca. T47D	13.6	Skeletal Muscle Pool	1.0
Breast ca. MDA-N	0.0	Spleen Pool	0.5
Breast Pool	12.3	Thymus Pool	1.5
Trachea	3.0	CNS cancer (glio/astro) U87-MG	0.2
Lung	1.6	CNS cancer (glio/astro) U- 118-MG	0.1
Fetal Lung	4.6	CNS cancer (neuro;met) SK-N-AS	0.1
Lung ca. NCI-N417	0.8	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	6.9	CNS cancer (astro) SNB-75	2.7
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	0.1
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.1
Lung ca. A549	0.0	Brain (Amygdala) Pool	3.0
Lung ca. NCI-H526	0.1	Brain (cerebellum)	100.0
Lung ca. NCI-H23	0.6	Brain (fetal)	12.3
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.0
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	9.5

Lung ca. NCI-H522	0.8	Brain (Substantia nigra) Pool	8.1
Liver	0.0	Brain (Thalamus) Pool	9.0
Fetal Liver	0.0	Brain (whole)	9.5
Liver ca. HepG2	0.2	Spinal Cord Pool	2.7
Kidney Pool	12.7	Adrenal Gland	0.9
Fetal Kidney	0.6	Pituitary gland Pool	4.9
Renal ca. 786-0	0.0	Salivary Gland	3.0
Renal ca. A498	0.1	Thyroid (female)	0.6
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.1
Renal ca. UO-31	0.1	Pancreas Pool	11.3

Table BD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3369, Run 165296636	Tissue Name	Rel. Exp.(%) Ag3369, Run 165296636
Secondary Th1 act	0.8	HUVEC IL-1beta	6.9
Secondary Th2 act	2.9	HUVEC IFN gamma	100.0
Secondary Tr1 act	1.7	HUVEC TNF alpha + IFN gamma	8.3
Secondary Th1 rest	5.7	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	1.3	HUVEC IL-11	10.9
Secondary Tr1 rest	1.4	Lung Microvascular EC none	30.4
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	9.3
Primary Th2 act	1.6	Microvascular Dermal EC none	17.8
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	7.9
Primary Th1 rest	3.6	Bronchial epithelium TNFalpha + IL1beta	3.8
Primary Th2 rest	7.4	Small airway epithelium none	0.8
Primary Tr1 rest	3.3	Small airway epithelium TNFalpha + IL-1beta	4.0
CD45RA CD4 lymphocyte act	4.8	Coronary artery SMC rest	8.7
CD45RO CD4 lymphocyte act	3.3	Coronary artery SMC TNFalpha + IL-1beta	1.4
CD8 lymphocyte act	3.8	Astrocytes rest	1.5
Secondary CD8 lymphocyte rest	1.7	Astrocytes TNFalpha + IL-1beta	11.9
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	3.0	CCD1106 (Keratinocytes) none	42.6
LAK cells rest	2.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	27.5
LAK cells IL-2	0.0	Liver cirrhosis	7.3
LAK cells IL-2+IL-12	0.9	Lupus kidney	4.6
LAK cells IL-2+IFN	9.5	NCI-H292 none	7.1

gamma			
LAK cells IL-2+ IL-18	4.0	NCI-H292 IL-4	5.9
LAK cells PMA/ionomycin	3.8	NCI-H292 IL-9	3.6
NK Cells IL-2 rest	1.7	NCI-H292 IL-13	8.5
Two Way MLR 3 day	14.5	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	6.5	HPAEC none	13.1
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	4.4
PBMC rest	4.2	Lung fibroblast none	4.4
PBMC PWM	14.2	Lung fibroblast TNF alpha + IL-1 beta	4.5
PBMC PHA-L	4.0	Lung fibroblast IL-4	7.9
Ramos (B cell) none	0.0	Lung fibroblast IL-9	9.2
Ramos (B cell) ionomycin	3.3	Lung fibroblast IL-13	3.6
B lymphocytes PWM	42.9	Lung fibroblast IFN gamma	28.5
B lymphocytes CD40L and IL-4	4.2	Dermal fibroblast CCD1070 rest	52.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	9.9
EOL-1 dbcAMP PMA/ionomycin	0.9	Dermal fibroblast CCD1070 IL-1 beta	12.3
Dendritic cells none	2.0	Dermal fibroblast IFN gamma	4.6
Dendritic cells LPS	15.0	Dermal fibroblast IL-4	3.5
Dendritic cells anti-CD40	1.8	IBD Colitis 2	2.8
Monocytes rest	1.6	IBD Crohn's	5.0
Monocytes LPS	6.6	Colon	65.1
Macrophages rest	2.9	Lung	88.3
Macrophages LPS	2.2	Thymus	6.7
HUVEC none	7.9	Kidney	24.5
HUVEC starved	18.7		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3369 This panel confirms the expression of the NOV2a gene at moderate levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between

5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3369 Expression of the NOV2a gene is highest in the cerebellum (CT=26.2). Therefore, expression of this gene can be used to

10 distinguish this sample from the others on the panel. In addition, this gene is expressed at moderate levels in hippocampus, thalamus, substantia nigra, cerebral cortex and spinal cord. The NOV2a gene encodes a protein with homology to rat neurexophilin 3. Neurexophilins are members of a family of neuropeptide-like glycoproteins that bind to alpha-Neurexins,

receptor-like proteins expressed on the neuronal cell surface (ref. 1). Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, pituitary gland, heart, and the gastrointestinal tract and at low levels in adrenal gland, thyroid, and skeletal muscle. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes. Expression of this gene is also significantly higher in adult heart (CT = 30) when compared to fetal heart (CT = 33.3), suggesting that it can be used to distinguish adult and fetal sources of this tissue.

Expression of this gene appears to be primarily associated with normal tissues rather than cancer cell lines. NOV2a gene expression appears to be down-regulated in CNS, colon, gastric, and renal cancer cell lines when compared to the corresponding normal tissues. Thus, expression of this gene may be useful as a marker for these types of cancers. Furthermore, application of the NOV2a gene product as a protein therapeutic may be of benefit in the treatment of CNS, colon, gastric, and renal cancer (Missler and Sudhof 1998)).

**Panel 4D Summary:** Ag3369 Highest expression of the NOV2a gene is seen in gamma interferon treated HUVECs (CT=31.6). Therefore, regulation of the transcript expression in HUVECs suggests that the protein encoded by this transcript may contribute to the inflammatory changes due to gamma interferon. Therefore, therapies designed with the protein encoded by this transcript may reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which endothelial cells and astrocytes are involved, such as lupus erythematosus, asthma, emphysema, Crohn's disease, ulcerative colitis, multiple sclerosis, rheumatoid arthritis, osteoarthritis, and psoriasis.

Significant levels of expression are also seen in normal colon and lung, suggesting that therapeutic modulation of the activity of this protein may be useful in the treatment of inflammatory bowel and lung diseases.

### C. NOV3a: PROTEASE INHIBITOR 9

Expression of gene NOV3a was assessed using the primer-probe set Ag3368, described in Table CA.

Table CA. Probe Name Ag3368

Primers	Sequences	Length	Start Position
---------	-----------	--------	----------------

Forward	5'-gacgagaccactgacttgagaa-3' (SEQ ID NO:203)	22	728
Probe	TET-5'-tcacttttgagaaactcacagcctgg-3'-TAMRA (SEQ ID NO:204)	26	765
Reverse	5'-tcttcatacagctctggcttgg-3' (SEQ ID NO:205)	22	791

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3368 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**General\_screening\_panel\_v1.4 Summary:** Ag3368 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

- 5 **Panel 4D Summary:** Ag3368 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### D. NOV6a: Growth Suppressor/Leprecan

- 10 Expression of gene NOV6a was assessed using the primer-probe set Ag3354, described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB, DC and DD.

**Table DA.** Probe Name Ag3354

Primers	Sequences	Length	Start Position
Forward	5'-gcagcacacaccttctttgtag-3' (SEQ ID NO:206)	22	561
Probe	TET-5'-caaaccatgcacctgcagatg-3'-TAMRA (SEQ ID NO:207)	23	583
Reverse	5'-ccgacattcgtctgtacttagc-3' (SEQ ID NO:208)	22	618

**Table DB.** CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3354, Run 206533686	Tissue Name	Rel. Exp.(%) Ag3354, Run 206533686
AD 1 Hippo	21.2	Control (Path) 3 Temporal Ctx	7.6
AD 2 Hippo	25.0	Control (Path) 4 Temporal Ctx	36.9
AD 3 Hippo	8.6	AD 1 Occipital Ctx	34.6
AD 4 Hippo	6.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	11.9
AD 6 Hippo	31.0	AD 4 Occipital Ctx	17.6
Control 2 Hippo	4.7	AD 5 Occipital Ctx	15.5
Control 4 Hippo	12.0	AD 6 Occipital Ctx	16.8
Control (Path) 3 Hippo	5.8	Control 1 Occipital Ctx	3.0
AD 1 Temporal Ctx	26.6	Control 2 Occipital Ctx	32.8
AD 2 Temporal Ctx	25.2	Control 3 Occipital Ctx	37.1
AD 3 Temporal Ctx	11.4	Control 4 Occipital Ctx	6.0
AD 4 Temporal Ctx	23.3	Control (Path) 1 Occipital Ctx	50.0
AD 5 Inf Temporal Ctx	55.1	Control (Path) 2 Occipital Ctx	16.5
AD 5 SupTemporal Ctx	41.8	Control (Path) 3 Occipital Ctx	2.0

AD 6 Inf Temporal Ctx	39.0	Control (Path) 4 Occipital Ctx	35.1
AD 6 Sup Temporal Ctx	44.8	Control 1 Parietal Ctx	11.8
Control 1 Temporal Ctx	5.8	Control 2 Parietal Ctx	59.0
Control 2 Temporal Ctx	22.2	Control 3 Parietal Ctx	16.4
Control 3 Temporal Ctx	23.7	Control (Path) 1 Parietal Ctx	39.2
Control 4 Temporal Ctx	17.8	Control (Path) 2 Parietal Ctx	31.9
Control (Path) 1 Temporal Ctx	39.2	Control (Path) 3 Parietal Ctx	4.7
Control (Path) 2 Temporal Ctx	32.3	Control (Path) 4 Parietal Ctx	35.4

Table DC. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3354, Run 174285052	Tissue Name	Rel. Exp.(%) Ag3354, Run 174285052
Normal Colon	5.1	Kidney Margin (OD04348)	32.8
Colon cancer (OD06064)	13.8	Kidney malignant cancer (OD06204B)	7.1
Colon Margin (OD06064)	2.5	Kidney normal adjacent tissue (OD06204E)	4.4
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	16.0
Colon Margin (OD06159)	16.8	Kidney Margin (OD04450-03)	7.2
Colon cancer (OD06297-04)	1.9	Kidney Cancer 8120613	1.5
Colon Margin (OD06297-05)	6.2	Kidney Margin 8120614	1.5
CC Gr.2 ascend colon (ODO3921)	9.7	Kidney Cancer 9010320	10.2
CC Margin (ODO3921)	6.2	Kidney Margin 9010321	3.6
Colon cancer metastasis (OD06104)	3.2	Kidney Cancer 8120607	34.4
Lung Margin (OD06104)	1.9	Kidney Margin 8120608	4.5
Colon mets to lung (OD04451-01)	13.5	Normal Uterus	34.6
Lung Margin (OD04451-02)	6.2	Uterine Cancer 064011	7.1
Normal Prostate	7.4	Normal Thyroid	5.6
Prostate Cancer (OD04410)	5.4	Thyroid Cancer 064010	15.5
Prostate Margin (OD04410)	6.8	Thyroid Cancer A302152	59.0
Normal Ovary	100.0	Thyroid Margin A302153	3.5
Ovarian cancer (OD06283-03)	14.2	Normal Breast	4.5
Ovarian Margin (OD06283-07)	7.6	Breast Cancer (OD04566)	6.6
Ovarian Cancer 064008	11.7	Breast Cancer 1024	7.4
Ovarian cancer (OD06145)	11.0	Breast Cancer (OD04590-01)	17.2
Ovarian Margin (OD06145)	13.0	Breast Cancer Mets (OD04590-03)	6.9



Ovarian cancer (OD06455-03)	5.4	Breast Cancer Metastasis (OD04655-05)	0.0
Ovarian Margin (OD06455-07)	7.3	Breast Cancer 064006	17.7
Normal Lung	7.8	Breast Cancer 9100266	6.7
Invasive poor diff. lung adeno (ODO4945-01)	11.6	Breast Margin 9100265	11.9
Lung Margin (ODO4945-03)	4.8	Breast Cancer A209073	3.4
Lung Malignant Cancer (OD03126)	19.3	Breast Margin A2090734	22.8
Lung Margin (OD03126)	1.6	Breast cancer (OD06083)	19.9
Lung Cancer (OD05014A)	5.1	Breast cancer node metastasis (OD06083)	39.5
Lung Margin (OD05014B)	8.3	Normal Liver	2.8
Lung cancer (OD06081)	2.7	Liver Cancer 1026	7.6
Lung Margin (OD06081)	2.3	Liver Cancer 1025	3.7
Lung Cancer (OD04237-01)	6.3	Liver Cancer 6004-T	1.3
Lung Margin (OD04237-02)	19.9	Liver Tissue 6004-N	4.8
Ocular Melanoma Metastasis	1.5	Liver Cancer 6005-T	15.0
Ocular Melanoma Margin (Liver)	2.8	Liver Tissue 6005-N	20.2
Melanoma Metastasis	12.2	Liver Cancer 064003	0.6
Melanoma Margin (Lung)	2.7	Normal Bladder	12.7
Normal Kidney	1.8	Bladder Cancer 1023	13.9
Kidney Ca, Nuclear grade 2 (OD04338)	14.1	Bladder Cancer A302173	8.8
Kidney Margin (OD04338)	9.2	Normal Stomach	21.5
Kidney Ca Nuclear grade 1/2 (OD04339)	6.7	Gastric Cancer 9060397	5.3
Kidney Margin (OD04339)	0.6	Stomach Margin 9060396	6.7
Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer 9060395	11.6
Kidney Margin (OD04340)	3.6	Stomach Margin 9060394	13.2
Kidney Ca, Nuclear grade 3 (OD04348)	38.7	Gastric Cancer 064005	7.6

Table DD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3354, Run 165241958	Tissue Name	Rel. Exp.(%) Ag3354, Run 165241958
Secondary Th1 act	0.2	HUVEC IL-1beta	10.6
Secondary Th2 act	0.4	HUVEC IFN gamma	42.0
Secondary Tr1 act	0.8	HUVEC TNF alpha + IFN gamma	28.7
Secondary Th1 rest	0.8	HUVEC TNF alpha + IL4	39.8
Secondary Th2 rest	0.4	HUVEC IL-11	44.1
Secondary Tr1 rest	0.1	Lung Microvascular EC none	61.6
Primary Th1 act	0.7	Lung Microvascular EC TNFalpha + IL-1beta	49.7
Primary Th2 act	0.9	Microvascular Dermal EC none	40.3

Primary Tr1 act	0.6	Microvascular Dermal EC TNFalpha + IL-1beta	22.7
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	35.8
Primary Th2 rest	0.2	Small airway epithelium none	5.2
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	4.2
CD45RA CD4 lymphocyte act	17.3	Coronary artery SMC rest	39.8
CD45RO CD4 lymphocyte act	0.3	Coronary artery SMC TNFalpha + IL-1beta	32.8
CD8 lymphocyte act	0.1	Astrocytes rest	27.2
Secondary CD8 lymphocyte rest	0.8	Astrocytes TNFalpha + IL-1beta	22.5
Secondary CD8 lymphocyte act	0.4	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	1.0	KU-812 (Basophil) PMA/ionomycin	0.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.2	CCD1106 (Keratinocytes) none	19.9
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.0
LAK cells IL-2	0.6	Liver cirrhosis	3.0
LAK cells IL-2+IL-12	0.4	Lupus kidney	1.2
LAK cells IL-2+IFN gamma	0.3	NCI-H292 none	1.2
LAK cells IL-2+ IL-18	0.6	NCI-H292 IL-4	0.2
LAK cells PMA/ionomycin	0.1	NCI-H292 IL-9	0.7
NK Cells IL-2 rest	0.4	NCI-H292 IL-13	0.2
Two Way MLR 3 day	0.3	NCI-H292 IFN gamma	0.1
Two Way MLR 5 day	0.0	HPAEC none	60.7
Two Way MLR 7 day	0.6	HPAEC TNF alpha + IL-1 beta	39.5
PBMC rest	0.3	Lung fibroblast none	70.7
PBMC PWM	0.2	Lung fibroblast TNF alpha + IL- 1 beta	54.0
PBMC PHA-L	0.3	Lung fibroblast IL-4	94.6
Ramos (B cell) none	0.0	Lung fibroblast IL-9	76.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	62.9
B lymphocytes PWM	0.9	Lung fibroblast IFN gamma	80.7
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	100.0
EOL-1 dbcAMP	0.9	Dermal fibroblast CCD1070 TNF alpha	68.3
EOL-1 dbcAMP PMA/ionomycin	0.2	Dermal fibroblast CCD1070 IL- 1 beta	63.7
Dendritic cells none	0.9	Dermal fibroblast IFN gamma	18.3
Dendritic cells LPS	0.1	Dermal fibroblast IL-4	46.7
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.5
Monocytes rest	1.1	IBD Crohn's	3.1
Monocytes LPS	0.3	Colon	10.5

Macrophages rest	0.2	Lung	17.2
Macrophages LPS	0.0	Thymus	6.3
HUVEC none	55.1	Kidney	4.5
HUVEC starved	76.3		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3354 This panel confirms the expression of the NOV6a gene at low to moderate levels in the brains of several individuals. However, no differential expression of this gene was detected between Alzheimer's diseased  
5 postmortem brains and those of non-demented controls in this experiment.

**General\_screening\_panel\_v1.4 Summary:** Ag3354 Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 2.2 Summary:** Ag3354 Highest expression of the NOV6a gene is seen in normal ovary (CT=32.3). Thus, expression of this gene could be used to differentiate between  
10 this sample and other samples on this panel and as a marker of ovarian tissue. The NOV6a gene encodes a protein with homology to the human Gros1 and rat leprecan genes. Stable transfection of the mouse Gros1 cDNA into NIH3T3 cells resulted in their slow growth and reduced colony-forming efficiency, suggesting that this protein can act as a growth suppressor. Therefore, use of the NOV6a gene product as a protein therapeutic may be of benefit in the  
15 treatment of cancer (Kaul et al., Oncogene 19(32):3576-83, 2000).

**Panel 4D Summary:** Ag3354 Expression of the NOV6a gene is highest in dermal and lung fibroblasts, regardless of treatment (CTs = 28-30). This gene is also expressed at moderate levels in endothelial cells. Thus, the transcript or the protein it encodes could be used to identify endothelium or fibroblasts. Endothelial cells are known to play important roles  
20 in inflammatory responses by altering the expression of surface proteins that are involved in activation and recruitment of effector inflammatory cells. The expression of this gene in dermal fibroblasts and dermal microvascular endothelial cells suggests that this protein product may be involved in inflammatory responses to skin disorders, including psoriasis. Expression in lung fibroblasts and lung microvascular endothelial cells suggests that the  
25 protein encoded by this transcript may also be involved in lung disorders including asthma, allergies, chronic obstructive pulmonary disease, and emphysema. Therefore, therapeutic modulation of the protein encoded by this gene may lead to amelioration of symptoms associated with psoriasis, asthma, allergies, chronic obstructive pulmonary disease, and emphysema.

30

**E. NOV10a: olfactomedin-like**

Expression of gene NOV10a was assessed using the primer-probe set Ag3384, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB, EC, ED, EE and EF.

5 **Table EA. Probe Name Ag3384**

Primers	Sequences	Length	Start Position
Forward	5'-actactatcggctgtgcaaatac-3' (SEQ ID NO:209)	22	826
Probe	TET-5'-ctataatgacctcgactgctgaaaa-3'-TAMRA (SEQ ID NO:210)	26	848
Reverse	5'-catagcccatcttcctctcttc-3' (SEQ ID NO:211)	22	879

**Table EB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag3384, Run 210154823	Tissue Name	Rel. Exp.(%) Ag3384, Run 210154823
AD 1 Hippo	36.6	Control (Path) 3 Temporal Ctx	6.8
AD 2 Hippo	48.0	Control (Path) 4 Temporal Ctx	43.2
AD 3 Hippo	5.3	AD 1 Occipital Ctx	39.2
AD 4 Hippo	3.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	50.0	AD 3 Occipital Ctx	0.0
AD 6 Hippo	62.0	AD 4 Occipital Ctx	27.5
Control 2 Hippo	7.6	AD 5 Occipital Ctx	13.0
Control 4 Hippo	32.1	AD 6 Occipital Ctx	18.7
Control (Path) 3 Hippo	69.7	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	55.5	Control 2 Occipital Ctx	7.2
AD 2 Temporal Ctx	36.3	Control 3 Occipital Ctx	26.1
AD 3 Temporal Ctx	14.1	Control 4 Occipital Ctx	3.2
AD 4 Temporal Ctx	86.5	Control (Path) 1 Occipital Ctx	36.6
AD 5 Inf Temporal Ctx	67.4	Control (Path) 2 Occipital Ctx	14.4
AD 5 Sup Temporal Ctx	100.0	Control (Path) 3 Occipital Ctx	40.9
AD 6 Inf Temporal Ctx	40.6	Control (Path) 4 Occipital Ctx	14.4
AD 6 Sup Temporal Ctx	54.3	Control 1 Parietal Ctx	27.2
Control 1 Temporal Ctx	6.0	Control 2 Parietal Ctx	95.9
Control 2 Temporal Ctx	10.8	Control 3 Parietal Ctx	28.1
Control 3 Temporal Ctx	26.1	Control (Path) 1 Parietal Ctx	20.6
Control 4 Temporal Ctx	6.3	Control (Path) 2 Parietal Ctx	33.7
Control (Path) 1 Temporal Ctx	14.4	Control (Path) 3 Parietal Ctx	26.6
Control (Path) 2	29.7	Control (Path) 4	55.5

Temporal Ctx		Parietal Ctx	
--------------	--	--------------	--

Table EC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3384, Run 213510091	Tissue Name	Rel. Exp.(%) Ag3384, Run 213510091
Adipose	4.1	Renal ca. TK-10	1.4
Melanoma* Hs688(A).T	1.8	Bladder	29.3
Melanoma* Hs688(B).T	4.7	Gastric ca. (liver met.) NCI-N87	61.1
Melanoma* M14	1.0	Gastric ca. KATO III	0.7
Melanoma* LOXIMVI	3.4	Colon ca. SW-948	0.9
Melanoma* SK-MEL-5	0.3	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.4	Colon ca.* (SW480 met) SW620	0.4
Testis Pool	14.2	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	3.6	Colon ca. HCT-116	1.6
Prostate Pool	10.3	Colon ca. CaCo-2	1.5
Placenta	7.5	Colon cancer tissue	0.5
Uterus Pool	4.4	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	35.1	Colon ca. Colo-205	0.7
Ovarian ca. SK-OV-3	100.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.9	Colon Pool	26.2
Ovarian ca. OVCAR-5	8.6	Small Intestine Pool	31.2
Ovarian ca. IGROV-1	0.0	Stomach Pool	25.9
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	13.6
Ovary	24.0	Fetal Heart	18.9
Breast ca. MCF-7	0.0	Heart Pool	7.6
Breast ca. MDA-MB-231	0.7	Lymph Node Pool	34.9
Breast ca. BT 549	2.7	Fetal Skeletal Muscle	8.0
Breast ca. T47D	7.5	Skeletal Muscle Pool	2.6
Breast ca. MDA-N	0.0	Spleen Pool	23.2
Breast Pool	31.2	Thymus Pool	30.6
Trachea	14.0	CNS cancer (glio/astro) U87-MG	2.6
Lung	16.3	CNS cancer (glio/astro) U-118-MG	3.7
Fetal Lung	69.7	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.4	CNS cancer (astro) SF-539	1.3
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	3.7
Lung ca. NCI-H146	5.8	CNS cancer (glio) SNB-19	1.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	27.4
Lung ca. A549	1.9	Brain (Amygdala) Pool	2.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.4
Lung ca. NCI-H23	13.8	Brain (fetal)	9.9
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.6

Lung ca. HOP-62	11.4	Cerebral Cortex Pool	2.0
Lung ca. NCI-H522	0.9	Brain (Substantia nigra) Pool	4.0
Liver	0.6	Brain (Thalamus) Pool	2.5
Fetal Liver	11.2	Brain (whole)	3.5
Liver ca. HepG2	0.9	Spinal Cord Pool	7.9
Kidney Pool	42.9	Adrenal Gland	10.9
Fetal Kidney	43.8	Pituitary gland Pool	3.5
Renal ca. 786-0	3.1	Salivary Gland	5.4
Renal ca. A498	3.3	Thyroid (female)	4.3
Renal ca. ACHN	4.5	Pancreatic ca. CAPAN2	3.0
Renal ca. UO-31	11.3	Pancreas Pool	29.1

Table ED. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3384, Run 173761690	Tissue Name	Rel. Exp.(%) Ag3384, Run 173761690
Normal Colon	8.2	Kidney Margin (OD04348)	100.0
Colon cancer (OD06064)	0.0	Kidney malignant cancer (OD06204B)	3.2
Colon Margin (OD06064)	3.3	Kidney normal adjacent tissue (OD06204E)	0.8
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	6.0
Colon Margin (OD06159)	7.5	Kidney Margin (OD04450-03)	14.2
Colon cancer (OD06297-04)	0.0	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	9.0	Kidney Margin 8120614	2.2
CC Gr.2 ascend colon (ODO3921)	4.3	Kidney Cancer 9010320	2.8
CC Margin (ODO3921)	0.0	Kidney Margin 9010321	7.3
Colon cancer metastasis (OD06104)	4.8	Kidney Cancer 8120607	0.0
Lung Margin (OD06104)	4.2	Kidney Margin 8120608	0.0
Colon mets to lung (OD04451-01)	4.7	Normal Uterus	28.1
Lung Margin (OD04451-02)	6.8	Uterine Cancer 064011	2.3
Normal Prostate	7.2	Normal Thyroid	2.1
Prostate Cancer (OD04410)	6.8	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	13.3	Thyroid Cancer A302152	13.7
Normal Ovary	0.0	Thyroid Margin A302153	0.0
Ovarian cancer (OD06283-03)	2.2	Normal Breast	17.3
Ovarian Margin (OD06283-07)	3.8	Breast Cancer (OD04566)	2.5
Ovarian Cancer 064008	26.1	Breast Cancer 1024	5.1
Ovarian cancer (OD06145)	4.5	Breast Cancer (OD04590-01)	3.7
Ovarian Margin (OD06145)	7.2	Breast Cancer Mets (OD04590-03)	7.6

Ovarian cancer (OD06455-03)	6.9	Breast Cancer Metastasis (OD04655-05)	2.2
Ovarian Margin (OD06455-07)	0.0	Breast Cancer 064006	9.5
Normal Lung	2.4	Breast Cancer 9100266	2.6
Invasive poor diff. lung adeno (ODO4945-01)	17.8	Breast Margin 9100265	1.8
Lung Margin (ODO4945-03)	10.2	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	1.2
Lung Margin (OD03126)	4.7	Breast cancer (OD06083)	15.8
Lung Cancer (OD05014A)	2.7	Breast cancer node metastasis (OD06083)	9.1
Lung Margin (OD05014B)	7.4	Normal Liver	16.5
Lung cancer (OD06081)	0.0	Liver Cancer 1026	2.0
Lung Margin (OD06081)	9.9	Liver Cancer 1025	13.8
Lung Cancer (OD04237-01)	2.4	Liver Cancer 6004-T	1.3
Lung Margin (OD04237-02)	21.9	Liver Tissue 6004-N	2.1
Ocular Melanoma Metastasis	2.3	Liver Cancer 6005-T	0.0
Ocular Melanoma Margin (Liver)	0.0	Liver Tissue 6005-N	3.0
Melanoma Metastasis	0.0	Liver Cancer 064003	0.0
Melanoma Margin (Lung)	9.1	Normal Bladder	4.7
Normal Kidney	4.0	Bladder Cancer 1023	0.7
Kidney Ca, Nuclear grade 2 (OD04338)	30.4	Bladder Cancer A302173	19.8
Kidney Margin (OD04338)	16.5	Normal Stomach	28.7
Kidney Ca Nuclear grade 1/2 (OD04339)	25.9	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	8.4	Stomach Margin 9060396	5.6
Kidney Ca, Clear cell type (OD04340)	2.3	Gastric Cancer 9060395	6.8
Kidney Margin (OD04340)	5.2	Stomach Margin 9060394	2.3
Kidney Ca, Nuclear grade 3 (OD04348)	5.0	Gastric Cancer 064005	0.0

Table EE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3384, Run 165296536	Tissue Name	Rel. Exp.(%) Ag3384, Run 165296536
Secondary Th1 act	0.0	HUVEC IL-1beta	2.2
Secondary Th2 act	2.9	HUVEC IFN gamma	7.8
Secondary Tr1 act	13.0	HUVEC TNF alpha + IFN gamma	3.4
Secondary Th1 rest	10.1	HUVEC TNF alpha + IL4	2.7
Secondary Th2 rest	10.0	HUVEC IL-11	1.2
Secondary Tr1 rest	6.0	Lung Microvascular EC none	15.0
Primary Th1 act	2.0	Lung Microvascular EC TNFalpha + IL-1beta	7.7
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0

Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	2.6
Primary Th1 rest	56.3	Bronchial epithelium TNFalpha + IL1beta	8.7
Primary Th2 rest	26.1	Small airway epithelium none	5.5
Primary Tr1 rest	4.1	Small airway epithelium TNFalpha + IL-1beta	50.7
CD45RA CD4 lymphocyte act	4.9	Coronary artery SMC rest	8.2
CD45RO CD4 lymphocyte act	7.1	Coronary artery SMC TNFalpha + IL-1beta	4.4
CD8 lymphocyte act	8.5	Astrocytes rest	9.1
Secondary CD8 lymphocyte rest	8.9	Astrocytes TNFalpha + IL-1beta	8.2
Secondary CD8 lymphocyte act	7.0	KU-812 (Basophil) rest	2.5
CD4 lymphocyte none	8.5	KU-812 (Basophil) PMA/ionomycin	33.9
2ry Th1/Th2/Tr1_anti- CD95 CH11	7.3	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	13.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.3
LAK cells IL-2	30.6	Liver cirrhosis	20.0
LAK cells IL-2+IL-12	25.7	Lupus kidney	11.3
LAK cells IL-2+IFN gamma	20.0	NCI-H292 none	54.7
LAK cells IL-2+ IL-18	28.1	NCI-H292 IL-4	30.4
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	39.0
NK Cells IL-2 rest	5.0	NCI-H292 IL-13	15.4
Two Way MLR 3 day	27.0	NCI-H292 IFN gamma	17.3
Two Way MLR 5 day	5.0	HPAEC none	13.7
Two Way MLR 7 day	7.7	HPAEC TNF alpha + IL-1 beta	2.4
PBMC rest	4.2	Lung fibroblast none	24.5
PBMC PWM	27.0	Lung fibroblast TNF alpha + IL- 1 beta	12.5
PBMC PHA-L	1.8	Lung fibroblast IL-4	15.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	20.9
Ramos (B cell) ionomycin	5.9	Lung fibroblast IL-13	2.6
B lymphocytes PWM	7.7	Lung fibroblast IFN gamma	18.6
B lymphocytes CD40L and IL-4	1.4	Dermal fibroblast CCD1070 rest	7.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	15.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL- 1 beta	0.0
Dendritic cells none	3.0	Dermal fibroblast IFN gamma	13.6
Dendritic cells LPS	4.4	Dermal fibroblast IL-4	5.4
Dendritic cells anti-CD40	3.4	IBD Colitis 2	2.4
Monocytes rest	7.3	IBD Crohn's	0.0
Monocytes LPS	2.7	Colon	11.2



Macrophages rest	4.5	Lung	8.0
Macrophages LPS	0.0	Thymus	88.3
HUVEC none	2.1	Kidney	100.0
HUVEC starved	26.1		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3384 The NOV10a gene, an olfactomedin homolog, is slightly upregulated in the temporal cortex of Alzheimer's disease patients. Members of the olfactomedin family have been implicated in regulating physical properties of the extracellular environment. Therefore, therapeutic inhibition of this protein may be of use in reversing the dementia/memory loss associated with Alzheimer's disease and neuronal death (Kulkarni et al., Genet Res 76(1):41-50, 2000).

**General screening panel\_v1.4 Summary:** Ag3384 Expression of the NOV10a gene is highest in an ovarian cancer cell line (CT=30.4). Significant expression of this gene is also seen in a gastric cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of ovarian and gastric cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of ovarian and gastric cancer.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, fetal skeletal muscle, fetal liver and adult/fetal heart. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

In addition, this gene has low expression in some samples derived from the central nervous system, including the substantia nigra, fetal brain, and spinal cord. Please see CNS\_neurodegeneration\_v1.0 for further discussion of the utility of this gene in the central nervous system.

**Panel 2.2 Summary:** Ag3384 In agreement with Panel 4D below, this gene is expressed at significant levels in the kidney, with highest expression in the kidney margin sample OD04348 (CT = 32.6). There is also low expression in samples from stomach, uterus, lung, and ovary. Thus, expression of this gene could be used to differentiate between the kidney and other samples on this panel and as a marker for kidney tissue.

**Panel 4D Summary:** Ag3384 Expression of the NOV10a gene is highest in kidney (CT=32.5) and thymus (CT=32.7). Therefore, protein, antibody or small molecule therapies designed with the NOV10a protein could be used to modulate kidney or T cell development

and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney and thymus, including lupus, glomerulonephritis, organ transplant, AIDS treatment or post chemotherapy immune reconstitution.

**Panel CNS\_1 Summary:** Ag3384 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

#### **F. NOV12a and NOV13a: NEURAL CELL ADHESION PROTEIN BIG-2 PRECURSOR**

Expression of genes NOV12a and NOV13a was assessed using the primer-probe sets Ag3228, Ag3261, Ag5267 and Ag5268, described in Tables FA, FB, FC and FD. Results of the RTQ-PCR runs are shown in Tables FE, FF, FG, FH, FI, FJ and FK.

**Table FA. Probe Name Ag3228**

Primers	Sequences	Length	Start Position
Forward	5'-gcccttccaagtttactga-3' (SEQ ID NO:212)	21	3266
Probe	TET-5'-tccttttaccctcatgctatccctga-3'-TAMRA (SEQ ID NO:213)	26	3291
Reverse	5'-gtaacgtgggcattattgacat-3' (SEQ ID NO:214)	22	3321

**Table FB. Probe Name Ag3261**

Primers	Sequences	Length	Start Position
Forward	5'-gcccttccaagtttactga-3' (SEQ ID NO:215)	21	3266
Probe	TET-5'-tccttttaccctcatgctatccctga-3'-TAMRA (SEQ ID NO:216)	26	3291
Reverse	5'-gtaacgtgggcattattgacat-3' (SEQ ID NO:217)	22	3321

**Table FC. Probe Name Ag5267**

Primers	Sequences	Length	Start Position
Forward	5'-gcggtcccgaaca-3' (SEQ ID NO:218)	14	2810
Probe	TET-5'-cacgcctggtctctcagtggca-3'-TAMRA (SEQ ID NO:219)	22	2841
Reverse	5'-gcctgctgccacacatt-3' (SEQ ID NO:220)	17	2871

**Table FD. Probe Name Ag5268**

Primers	Sequences	Length	Start Position
Forward	5'-cagcatcccttcagtga-3' (SEQ ID NO:221)	18	1013
Probe	TET-5'-cacagccaccaacaatgtgggc-3'-TAMRA (SEQ ID NO:222)	22	1058
Reverse	5'-caccagcaggttgacagtct-3' (SEQ ID NO:223)	20	1093

**Table FE. AI\_comprehensive panel\_v1.0**

Tissue Name	Rel. Exp.(%)	Rel. Exp.(%)	Rel. Exp.(%)	Rel. Exp.(%)	Rel. Exp.(%)
-------------	--------------	--------------	--------------	--------------	--------------

	Ag3228, Run 225147544	Ag3228, Run 229440553	Ag3261, Run 229313855	Ag5267, Run 230473002	Ag5268, Run 230473019
I10967 COPD-F	0.0	0.0	0.0	36.9	35.4
I10980 COPD-F	0.0	20.2	0.6	40.3	36.6
I10968 COPD-M	40.9	27.0	0.0	36.1	35.6
I10977 COPD-M	33.7	96.6	2.6	0.0	0.0
I10989 Emphysema-F	10.6	25.0	0.0	25.2	29.5
I10992 Emphysema-F	0.0	0.0	0.0	12.2	17.9
I10993 Emphysema-F	0.0	0.0	0.0	15.2	18.4
I10994 Emphysema-F	0.0	0.0	0.0	11.1	23.0
I10995 Emphysema-F	0.0	24.3	0.0	18.9	36.9
I10996 Emphysema-F	0.0	0.0	0.0	0.0	1.4
I10997 Asthma-M	0.0	0.0	0.0	21.2	7.0
I11001 Asthma-F	0.0	0.0	0.0	20.4	37.9
I11002 Asthma-F	9.7	0.0	0.0	22.4	19.3
I11003 Atopic Asthma-F	0.0	0.0	0.7	49.0	45.1
I11004 Atopic Asthma-F	0.0	0.0	0.0	49.3	59.9
I11005 Atopic Asthma-F	0.0	0.0	0.0	24.3	21.6
I11006 Atopic Asthma-F	0.0	0.0	0.0	6.5	5.6
I11417 Allergy-M	0.0	0.0	0.0	26.4	36.6
I12347 Allergy-M	8.5	0.0	1.0	1.5	3.4
I12349 Normal Lung-F	0.0	45.4	1.2	2.9	10.3
I12357 Normal Lung-F	20.2	77.9	1.0	51.4	38.7
I12354 Normal Lung-M	0.0	23.7	0.0	39.8	21.3
I12374 Crohns-F	10.7	0.0	0.0	25.0	19.9
I12389 Match Control Crohns-F	12.3	0.0	0.5	18.2	8.2
I12375 Crohns-F	0.0	0.0	0.0	24.0	21.2
I12732 Match Control Crohns-F	10.9	100.0	1.4	22.7	46.3
I12725 Crohns-M	0.0	0.0	0.5	3.1	1.8
I12387 Match Control Crohns-M	100.0	0.0	0.0	20.9	19.5
I12378 Crohns-M	49.3	43.5	0.8	2.9	9.8
I12390 Match Control Crohns-M	9.2	0.0	1.3	45.1	43.2
I12726 Crohns-M	9.3	14.0	0.0	50.0	51.4
I12731 Match Control Crohns-M	0.0	28.7	1.8	31.0	43.2

112380 Ulcer Col-F	0.0	0.0	0.0	20.6	15.5
112734 Match Control Ulcer Col-F	40.6	58.6	2.2	37.4	46.7
112384 Ulcer Col-F	0.0	0.0	0.0	41.5	25.7
112737 Match Control Ulcer Col-F	9.9	0.0	0.0	27.5	21.3
112386 Ulcer Col-F	0.0	0.0	0.3	25.2	15.9
112738 Match Control Ulcer Col-F	0.0	0.0	0.0	3.0	3.1
112381 Ulcer Col-M	0.0	0.0	0.0	4.5	24.0
112735 Match Control Ulcer Col-M	0.0	0.0	0.0	16.4	4.0
112382 Ulcer Col-M	12.2	23.8	0.5	18.7	16.2
112394 Match Control Ulcer Col-M	0.0	0.0	0.0	6.4	9.4
112383 Ulcer Col-M	13.1	23.8	0.7	16.6	9.5
112736 Match Control Ulcer Col-M	0.0	0.0	0.6	14.7	11.3
112423 Psoriasis-F	7.2	24.7	0.0	40.6	15.6
112427 Match Control Psoriasis-F	34.9	49.7	2.8	84.1	53.2
112418 Psoriasis-M	22.8	54.0	100.0	52.1	21.3
112723 Match Control Psoriasis-M	0.0	0.0	0.0	10.4	11.1
112419 Psoriasis-M	0.0	21.6	1.0	61.1	35.4
112424 Match Control Psoriasis-M	0.0	75.3	0.7	23.7	10.7
112420 Psoriasis-M	35.1	15.5	0.0	100.0	100.0
112425 Match Control Psoriasis-M	0.0	23.2	0.0	77.4	38.4
104689 (MF) OA Bone-Backus	0.0	0.0	1.0	16.4	25.7
104690 (MF) Adj "Normal" Bone-Backus	0.0	0.0	0.0	10.4	17.9
104691 (MF) OA	0.0	0.0	0.0	9.6	33.9

Synovium-Backus					
104692 (BA) OA Cartilage-Backus	0.0	0.0	0.0	2.7	9.9
104694 (BA) OA Bone-Backus	0.0	25.0	0.7	21.2	39.5
104695 (BA) Adj "Normal" Bone- Backus	0.0	0.0	0.0	11.4	23.5
104696 (BA) OA Synovium-Backus	0.0	0.0	0.6	31.0	43.5
104700 (SS) OA Bone-Backus	23.0	0.0	0.0	28.5	31.2
104701 (SS) Adj "Normal" Bone- Backus	0.0	0.0	0.0	38.2	70.7
104702 (SS) OA Synovium-Backus	0.0	17.3	0.0	65.5	89.5
117093 OA Cartilage Rep7	11.6	18.7	1.5	34.6	62.4
112672 OA Bone5	0.0	0.0	0.0	61.1	47.0
112673 OA Synovium5	0.0	0.0	0.0	20.3	21.8
112674 OA Synovial Fluid cells5	0.0	0.0	0.0	33.7	27.4
117100 OA Cartilage Rep14	0.0	0.0	0.0	16.3	20.2
112756 OA Bone9	0.0	0.0	0.0	5.1	2.8
112757 OA Synovium9	0.0	0.0	0.0	2.2	10.4
112758 OA Synovial Fluid Cells9	0.0	0.0	0.4	10.3	12.9
117125 RA Cartilage Rep2	0.0	25.7	0.0	34.9	41.8
113492 Bone2 RA	0.0	0.0	0.8	27.2	19.6
113493 Synovium2 RA	0.0	0.0	0.0	20.0	17.6
113494 Syn Fluid Cells RA	9.6	35.8	1.0	35.6	20.0
113499 Cartilage4 RA	0.0	0.0	0.0	52.1	31.0
113500 Bone4 RA	0.0	0.0	1.0	43.8	37.6
113501 Synovium4 RA	0.0	0.0	0.0	33.9	19.8
113502 Syn Fluid Cells4 RA	0.0	48.6	0.0	19.5	10.6
113495 Cartilage3 RA	12.6	0.0	0.6	15.5	17.8
113496 Bone3 RA	9.4	24.5	0.7	24.0	18.3
113497 Synovium3 RA	0.0	0.0	0.0	5.4	14.1
113498 Syn Fluid	0.0	0.0	1.0	23.3	21.2

Cells3 RA					
117106 Normal Cartilage Rep20	0.0	0.0	0.0	22.7	27.2
113663 Bone3 Normal	0.0	0.0	0.2	4.4	4.3
113664 Synovium3 Normal	0.0	0.0	0.5	0.7	4.5
113665 Syn Fluid Cells3 Normal	0.0	0.0	0.3	1.0	3.2
117107 Normal Cartilage Rep22	0.0	0.0	0.0	8.7	9.8
113667 Bone4 Normal	0.0	29.3	0.8	22.1	39.2
113668 Synovium4 Normal	0.0	81.8	1.2	40.1	26.8
113669 Syn Fluid Cells4 Normal	0.0	0.0	0.0	25.0	41.5

Table FF. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3228, Run 206533575	Rel. Exp.(%) Ag3261, Run 209990365	Rel. Exp.(%) Ag5267, Run 230510331	Rel. Exp.(%) Ag5268, Run 230510332
AD 1 Hippo	85.3	0.0	27.5	27.4
AD 2 Hippo	0.0	0.0	48.3	60.3
AD 3 Hippo	23.7	100.0	26.6	26.2
AD 4 Hippo	12.2	15.2	20.6	32.3
AD 5 Hippo	22.4	45.7	60.7	76.8
AD 6 Hippo	71.2	33.9	66.4	100.0
Control 2 Hippo	0.0	0.0	60.7	31.4
Control 4 Hippo	24.5	39.5	22.4	43.8
Control (Path) 3 Hippo	0.0	3.1	0.6	5.9
AD 1 Temporal Ctx	0.0	38.2	62.4	35.4
AD 2 Temporal Ctx	0.0	12.9	56.3	5.9
AD 3 Temporal Ctx	0.0	54.7	20.0	34.6
AD 4 Temporal Ctx	21.8	23.3	42.9	12.0
AD 5 Inf Temporal Ctx	86.5	28.5	66.0	60.3
AD 5 Sup Temporal Ctx	28.3	52.9	61.1	92.7
AD 6 Inf Temporal Ctx	100.0	96.6	37.9	56.3
AD 6 Sup Temporal Ctx	39.8	67.8	57.8	46.7
Control 1 Temporal Ctx	52.5	37.9	13.6	16.0
Control 2	0.0	3.2	36.6	29.7

Temporal Ctx				
Control 3 Temporal Ctx	0.0	10.2	15.9	13.3
Control 3 Temporal Ctx	0.0	16.6	20.7	26.1
Control (Path) 1 Temporal Ctx	54.3	1.7	81.2	51.8
Control (Path) 2 Temporal Ctx	0.0	0.0	30.4	25.3
Control (Path) 3 Temporal Ctx	0.0	0.0	13.1	12.9
Control (Path) 4 Temporal Ctx	10.8	16.5	36.3	22.1
AD 1 Occipital Ctx	10.7	0.0	26.8	31.0
AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0	0.0
AD 3 Occipital Ctx	0.0	9.9	16.4	30.1
AD 4 Occipital Ctx	0.0	1.4	33.9	15.3
AD 5 Occipital Ctx	66.0	0.0	50.0	14.8
AD 6 Occipital Ctx	0.0	1.8	11.9	25.5
Control 1 Occipital Ctx	0.0	25.7	4.7	4.8
Control 2 Occipital Ctx	0.0	6.1	38.4	27.0
Control 3 Occipital Ctx	0.0	6.8	19.2	15.5
Control 4 Occipital Ctx	0.0	2.4	12.6	22.5
Control (Path) 1 Occipital Ctx	50.7	17.7	100.0	66.9
Control (Path) 2 Occipital Ctx	0.0	8.9	9.7	7.1
Control (Path) 3 Occipital Ctx	0.0	0.0	7.6	5.0
Control (Path) 4 Occipital Ctx	0.0	26.8	18.4	20.6
Control 1 Parietal Ctx	0.0	72.2	13.3	39.0
Control 2 Parietal Ctx	30.8	19.2	60.3	67.4
Control 3 Parietal Ctx	0.0	8.5	15.2	16.2
Control (Path) 1 Parietal Ctx	22.7	12.2	80.1	48.6
Control (Path) 2 Parietal Ctx	22.1	15.8	37.6	11.0
Control (Path) 3 Parietal Ctx	0.0	0.0	8.0	15.7

Control (Path) 4 Parietal Ctx	0.0	37.1	44.1	38.7
----------------------------------	-----	------	------	------

Table FG. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3228, Run 213333208	Rel. Exp.(%) Ag3261, Run 216512991	Tissue Name	Rel. Exp.(%) Ag3228, Run 213333208	Rel. Exp.(%) Ag3261, Run 216512991
Adipose	0.0	0.0	Renal ca. TK-10	0.0	0.0
Melanoma* Hs688(A).T	0.0	0.0	Bladder	0.0	0.0
Melanoma* Hs688(B).T	0.8	0.0	Gastric ca. (liver met.) NCI-N87	1.0	0.0
Melanoma* M14	0.0	0.0	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMV1	0.0	0.0	Colon ca. SW-948	0.0	0.0
Melanoma* SK- MEL-5	0.0	0.0	Colon ca. SW480	0.0	0.0
Squamous cell carcinoma SCC- 4	0.0	0.0	Colon ca. * (SW480 met) SW620	0.0	0.0
Testis Pool	0.0	0.0	Colon ca. HT29	1.1	0.0
Prostate ca.* (bone met) PC-3	0.0	0.0	Colon ca. HCT-116	0.0	0.0
Prostate Pool	0.0	0.0	Colon ca. CaCo-2	0.0	0.0
Placenta	0.0	0.0	Colon cancer tissue	0.0	0.0
Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	1.0	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK- OV-3	9.6	0.1	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.0	0.0	Colon Pool	1.6	44.8
Ovarian ca. OVCAR-5	0.0	0.0	Small Intestine Pool	1.8	0.0
Ovarian ca. IGROV-1	0.0	0.0	Stomach Pool	0.0	0.0
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	0.0	0.0
Ovary	1.2	0.0	Fetal Heart	0.0	0.0
Breast ca. MCF- 7	0.0	0.0	Heart Pool	0.0	0.0
Breast ca. MDA- MB-231	4.4	0.0	Lymph Node Pool	2.4	0.0
Breast ca. BT 549	1.1	0.0	Fetal Skeletal Muscle	0.8	0.0
Breast ca. T47D	0.0	0.0	Skeletal Muscle Pool	0.0	0.0
Breast ca. MDA- N	3.5	0.0	Spleen Pool	0.0	0.0
Breast Pool	0.0	0.0	Thymus Pool	0.8	0.0



Trachea	0.9	0.0	CNS cancer (glio/astro) U87-MG	24.7	100.0
Lung	0.0	0.0	CNS cancer (glio/astro) U-118-MG	6.6	0.1
Fetal Lung	0.0	0.0	CNS cancer (neuro;met) SK-N-AS	0.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	3.6	0.0	CNS cancer (astro) SNB-75	3.8	0.0
Lung ca. NCI-H146	0.8	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	1.8	0.0
Lung ca. A549	0.0	0.0	Brain (Amygdala) Pool	0.8	0.0
Lung ca. NCI-H526	0.0	0.0	Brain (cerebellum)	100.0	1.2
Lung ca. NCI-H23	6.2	0.0	Brain (fetal)	24.8	0.4
Lung ca. NCI-H460	0.4	0.0	Brain (Hippocampus) Pool	0.0	0.0
Lung ca. HOP-62	0.7	0.0	Cerebral Cortex Pool	1.1	0.0
Lung ca. NCI-H522	6.0	0.1	Brain (Substantia nigra) Pool	2.0	0.0
Liver	0.6	0.0	Brain (Thalamus) Pool	2.0	0.0
Fetal Liver	0.0	0.0	Brain (whole)	1.9	0.1
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	0.0	0.0
Kidney Pool	3.7	0.0	Adrenal Gland	0.0	1.5
Fetal Kidney	1.4	0.0	Pituitary gland Pool	0.0	0.0
Renal ca. 786-0	0.0	0.0	Salivary Gland	0.0	0.0
Renal ca. A498	2.3	0.0	Thyroid (female)	0.0	42.6
Renal ca. ACHN	0.5	0.0	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	1.8	2.4

Table FH. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5267, Run 232936653	Rel. Exp.(%) Ag5267, Run 254397162	Rel. Exp.(%) Ag5268, Run 232936654	Tissue Name	Rel. Exp.(%) Ag5267, Run 232936653	Rel. Exp.(%) Ag5267, Run 254397162	Rel. Exp.(%) Ag5268, Run 232936654
Adipose	0.1	1.6	1.2	Renal ca. TK-10	0.3	3.0	3.7
Melanoma* Hs688(A).T	0.4	5.2	4.4	Bladder	0.1	0.6	0.8
Melanoma*	0.6	9.5	7.9	Gastric ca. (liver	0.0	0.4	0.4

Hs688(B).T				met.) NCI-N87			
Melanoma* M14	0.0	0.1	0.4	Gastric ca. KATO III	0.0	0.0	0.0
Melanoma* LOXIMVI	0.0	0.7	0.6	Colon ca. SW- 948	0.0	0.0	0.0
Melanoma* SK-MEL-5	0.6	8.5	10.8	Colon ca. SW480	0.2	3.1	3.0
Squamous cell carcinoma SCC-4	0.0	0.0	0.1	Colon ca.* (SW480 met) SW620	0.0	0.6	0.3
Testis Pool	0.1	1.1	0.6	Colon ca. HT29	0.0	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.1	0.0	Colon ca. HCT- 116	0.2	2.7	4.3
Prostate Pool	0.1	1.3	1.2	Colon ca. CaCo- 2	0.0	0.1	0.1
Placenta	0.0	0.2	0.2	Colon cancer tissue	0.1	1.2	1.1
Uterus Pool	0.2	2.7	1.5	Colon ca. SW1116	0.0	0.2	0.4
Ovarian ca. OVCAR-3	0.0	0.4	0.5	Colon ca. Colo- 205	0.0	0.0	0.0
Ovarian ca. SK-OV-3	0.5	6.7	8.0	Colon ca. SW- 48	0.0	0.0	0.0
Ovarian ca. OVCAR-4	0.1	1.3	2.8	Colon Pool	0.3	3.2	2.7
Ovarian ca. OVCAR-5	0.1	2.0	1.9	Small Intestine Pool	0.4	5.5	4.7
Ovarian ca. IGROV-1	0.0	0.4	0.3	Stomach Pool	0.2	1.7	2.4
Ovarian ca. OVCAR-8	0.1	0.6	2.1	Bone Marrow Pool	0.2	3.0	1.5
Ovary	0.2	2.9	2.1	Fetal Heart	0.1	1.1	1.2
Breast ca. MCF-7	0.0	0.0	0.1	Heart Pool	100.0	2.9	1.5
Breast ca. MDA-MB- 231	0.5	7.5	11.1	Lymph Node Pool	0.5	7.2	4.4
Breast ca. BT 549	0.9	12.1	12.5	Fetal Skeletal Muscle	0.3	3.6	3.4
Breast ca. T47D	0.2	2.5	2.2	Skeletal Muscle Pool	0.1	0.9	0.7
Breast ca. MDA-N	0.3	3.0	6.1	Spleen Pool	0.1	1.4	3.2
Breast Pool	0.5	5.2	3.4	Thymus Pool	0.2	2.7	2.5
Trachea	0.1	0.9	1.3	CNS cancer (glio/astro) U87-MG	0.1	1.7	1.2
Lung	0.2	1.7	1.3	CNS cancer (glio/astro) U- 118-MG	0.6	9.5	9.1
Fetal Lung	0.3	4.4	5.0	CNS cancer	0.0	0.4	0.4

				(neuro;met) SK-N-AS			
Lung ca. NCI-N417	0.0	0.6	1.9	CNS cancer (astro) SF-539	0.0	0.7	0.7
Lung ca. LX-1	0.3	4.4	1.4	CNS cancer (astro) SNB-75	2.6	37.9	44.1
Lung ca. NCI-H146	0.0	0.1	0.1	CNS cancer (glio) SNB-19	0.0	0.5	0.6
Lung ca. SHP-77	0.0	0.3	0.4	CNS cancer (glio) SF-295	0.3	2.9	3.4
Lung ca. A549	0.3	3.5	3.2	Brain (Amygdala) Pool	0.2	2.3	2.7
Lung ca. NCI-H526	0.0	0.2	0.2	Brain (cerebellum)	8.5	100.0	100.0
Lung ca. NCI-H23	0.1	0.3	0.3	Brain (fetal)	3.3	45.4	34.2
Lung ca. NCI-H460	0.1	0.2	1.8	Brain (Hippocampus) Pool	0.4	4.1	3.5
Lung ca. HOP-62	0.1	1.2	2.4	Cerebral Cortex Pool	0.3	3.1	3.3
Lung ca. NCI-H522	1.2	20.7	18.6	Brain (Substantia nigra) Pool	0.2	2.8	2.8
Liver	0.0	0.2	0.2	Brain (Thalamus) Pool	0.3	4.5	4.8
Fetal Liver	0.0	0.3	0.3	Brain (whole)	1.0	10.8	7.5
Liver ca. HepG2	0.0	0.0	0.0	Spinal Cord Pool	0.3	5.3	6.0
Kidney Pool	0.8	9.2	9.2	Adrenal Gland	0.2	3.5	3.2
Fetal Kidney	0.1	0.7	1.2	Pituitary gland Pool	0.1	1.5	1.7
Renal ca. 786-0	0.0	0.3	0.2	Salivary Gland	0.0	0.4	0.2
Renal ca. A498	0.5	6.3	6.9	Thyroid (female)	0.0	0.4	0.1
Renal ca. ACHN	0.4	5.6	6.7	Pancreatic ca. CAPAN2	0.0	0.0	0.0
Renal ca. UO-31	0.2	2.2	2.6	Pancreas Pool	0.2	3.0	2.4

Table FI. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3228, Run 173762591	Tissue Name	Rel. Exp.(%) Ag3228, Run 173762591
Normal Colon	5.3	Kidney Margin (OD04348)	0.0
Colon cancer (OD06064)	0.0	Kidney malignant cancer (OD06204B)	0.0
Colon Margin (OD06064)	0.0	Kidney normal adjacent tissue (OD06204E)	0.0
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	0.0

Colon Margin (OD06159)	0.0	Kidney Margin (OD04450-03)	0.0
Colon cancer (OD06297-04)	0.0	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	0.0	Kidney Margin 8120614	0.0
CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer 9010320	0.0
CC Margin (ODO3921)	0.0	Kidney Margin 9010321	0.0
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	0.0
Lung Margin (OD06104)	0.0	Kidney Margin 8120608	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Uterus	0.0
Lung Margin (OD04451-02)	0.0	Uterine Cancer 064011	0.0
Normal Prostate	0.0	Normal Thyroid	0.0
Prostate Cancer (OD04410)	0.0	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	0.0	Thyroid Cancer A302152	5.4
Normal Ovary	0.0	Thyroid Margin A302153	0.0
Ovarian cancer (OD06283-03)	0.0	Normal Breast	0.0
Ovarian Margin (OD06283-07)	14.4	Breast Cancer (OD04566)	0.0
Ovarian Cancer 064008	100.0	Breast Cancer 1024	0.0
Ovarian cancer (OD06145)	0.0	Breast Cancer (OD04590-01)	0.0
Ovarian Margin (OD06145)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Ovarian cancer (OD06455-03)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Ovarian Margin (OD06455-07)	0.0	Breast Cancer 064006	0.0
Normal Lung	0.0	Breast Cancer 9100266	0.0
Invasive poor diff. lung adeno (ODO4945-01)	0.0	Breast Margin 9100265	0.0
Lung Margin (ODO4945-03)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Breast cancer (OD06083)	0.0
Lung Cancer (OD05014A)	0.0	Breast cancer node metastasis (OD06083)	0.0
Lung Margin (OD05014B)	0.0	Normal Liver	0.0
Lung cancer (OD06081)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD06081)	11.7	Liver Cancer 1025	23.8
Lung Cancer (OD04237-01)	0.0	Liver Cancer 6004-T	0.0
Lung Margin (OD04237-02)	0.0	Liver Tissue 6004-N	0.0
Ocular Melanoma Metastasis	0.0	Liver Cancer 6005-T	0.0
Ocular Melanoma Margin (Liver)	0.0	Liver Tissue 6005-N	0.0
Melanoma Metastasis	0.0	Liver Cancer 064003	0.0
Melanoma Margin (Lung)	0.0	Normal Bladder	5.0
Normal Kidney	0.0	Bladder Cancer 1023	0.0
Kidney Ca, Nuclear grade 2	0.0	Bladder Cancer A302173	0.0

(OD04338)			
Kidney Margin (OD04338)	7.3	Normal Stomach	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	0.0	Stomach Margin 9060396	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer 9060395	17.3
Kidney Margin (OD04340)	0.0	Stomach Margin 9060394	0.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 064005	0.0

Table FJ. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5267, Run 230510063	Rel. Exp.(%) Ag5268, Run 230510184	Tissue Name	Rel. Exp.(%) Ag5267, Run 230510063	Rel. Exp.(%) Ag5268, Run 230510184
Secondary Th1 act	0.0	0.6	HUVEC IL-1beta	1.4	0.8
Secondary Th2 act	4.3	7.5	HUVEC IFN gamma	4.8	5.8
Secondary Tr1 act	3.4	2.9	HUVEC TNF alpha + IFN gamma	3.8	8.4
Secondary Th1 rest	0.6	1.0	HUVEC TNF alpha + IL4	0.4	0.3
Secondary Th2 rest	2.1	0.8	HUVEC IL-11	0.4	0.0
Secondary Tr1 rest	0.8	0.5	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.3	Lung Microvascular EC TNFalpha + IL- 1beta	0.4	2.7
Primary Th2 act	0.6	1.3	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.8	1.1	Microvascular Dermal EC TNFalpha + IL- 1beta	0.2	0.0
Primary Th1 rest	0.2	0.0	Bronchial epithelium TNFalpha + IL1beta	0.7	0.2
Primary Th2 rest	0.3	0.5	Small airway epithelium none	0.2	0.1
Primary Tr1 rest	0.2	1.4	Small airway epithelium TNFalpha + IL-1beta	1.6	1.2
CD45RA CD4 lymphocyte act	1.6	0.9	Coronary artery SMC rest	0.9	0.9
CD45RO CD4 lymphocyte act	3.4	1.5	Coronary artery SMC TNFalpha + IL-1beta	1.9	2.6
CD8 lymphocyte act	0.4	1.4	Astrocytes rest	8.8	11.2
Secondary CD8 lymphocyte rest	0.1	1.3	Astrocytes TNFalpha + IL-1beta	11.0	14.9
Secondary CD8 lymphocyte act	0.2	0.0	KU-812 (Basophil) rest	1.4	0.0
CD4 lymphocyte none	0.9	2.3	KU-812 (Basophil) PMA/ionomycin	9.3	0.6
2ry	1.8	4.1	CCD1106	4.8	0.9

Th1/Th2/Tr1_anti-CD95 CH11			(Keratinocytes) none		
LAK cells rest	6.2	7.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.2	1.8
LAK cells IL-2	3.3	2.3	Liver cirrhosis	2.0	0.9
LAK cells IL-2+IL-12	1.2	0.2	NCI-H292 none	0.4	0.5
LAK cells IL-2+IFN gamma	0.7	3.5	NCI-H292 IL-4	1.8	0.6
LAK cells IL-2+ IL-18	0.6	2.2	NCI-H292 IL-9	1.4	1.7
LAK cells PMA/ionomycin	41.8	26.4	NCI-H292 IL-13	2.8	2.5
NK Cells IL-2 rest	9.0	9.8	NCI-H292 IFN gamma	5.6	8.7
Two Way MLR 3 day	2.9	1.8	HPAEC none	0.2	0.6
Two Way MLR 5 day	3.5	0.7	HPAEC TNF alpha + IL-1 beta	4.9	4.1
Two Way MLR 7 day	2.7	1.4	Lung fibroblast none	46.7	41.5
PBMC rest	0.3	0.0	Lung fibroblast TNF alpha + IL-1 beta	9.7	7.3
PBMC PWM	0.0	0.0	Lung fibroblast IL-4	23.5	24.8
PBMC PHA-L	1.2	1.1	Lung fibroblast IL-9	32.5	36.9
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	18.6	14.4
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IFN gamma	73.7	82.9
B lymphocytes PWM	0.7	1.9	Dermal fibroblast CCD1070 rest	1.8	1.0
B lymphocytes CD40L and IL-4	7.6	4.1	Dermal fibroblast CCD1070 TNF alpha	0.3	3.5
EOL-1 dbcAMP	7.7	9.9	Dermal fibroblast CCD1070 IL-1 beta	2.0	2.1
EOL-1 dbcAMP PMA/ionomycin	100.0	100.0	Dermal fibroblast IFN gamma	11.7	13.3
Dendritic cells none	0.5	1.1	Dermal fibroblast IL-4	4.0	3.5
Dendritic cells LPS	0.4	0.0	Dermal Fibroblasts rest	10.1	3.5
Dendritic cells anti-CD40	0.4	0.2	Neutrophils TNFa+LPS	0.0	0.0
Monocytes rest	0.0	0.0	Neutrophils rest	0.7	0.4
Monocytes LPS	3.0	4.2	Colon	0.6	0.0
Macrophages rest	0.0	0.9	Lung	0.7	0.2
Macrophages LPS	0.7	0.8	Thymus	1.1	0.7
HUVEC none	0.7	0.0	Kidney	2.4	0.7
HUVEC starved	1.2	1.8			

Table FK. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3228, Run 164389698	Rel. Exp.(%) Ag3261, Run 164537293	Tissue Name	Rel. Exp.(%) Ag3228, Run 164389698	Rel. Exp.(%) Ag3261, Run 164537293
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	2.8	2.2	HUVEC IFN gamma	0.0	7.5
Secondary Tr1 act	0.0	2.8	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	2.3	0.0	Lung Microvascular EC none	1.7	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	0.0	0.0
Primary Th2 act	6.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microvascular Dermal EC TNFalpha + IL- 1beta	0.0	0.0
Primary Th1 rest	2.3	0.0	Bronchial epithelium TNFalpha + IL1beta	2.4	8.9
Primary Th2 rest	0.0	2.2	Small airway epithelium none	0.0	0.0
Primary Tr1 rest	2.6	4.3	Small airway epithelium TNFalpha + IL-1beta	2.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	1.3	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL-1beta	0.0	0.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	2.6	3.4
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	20.7	30.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0
LAK cells IL-2	0.0	3.0	Liver cirrhosis	9.3	7.2
LAK cells IL-2+IL- 12	0.0	0.0	Lupus kidney	0.0	0.0
LAK cells IL-2+IFN gamma	3.4	16.0	NCI-H292 none	0.0	0.0
LAK cells IL-2+ IL- 18	10.3	9.2	NCI-H292 IL-4	0.0	6.8
LAK cells PMA/ionomycin	100.0	63.7	NCI-H292 IL-9	0.0	0.0

NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	0.0	4.2
Two Way MLR 3 day	8.6	8.5	NCI-H292 IFN gamma	5.1	0.0
Two Way MLR 5 day	2.1	16.8	HPAEC none	0.0	0.0
Two Way MLR 7 day	14.7	2.3	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast none	2.4	0.0
PBMC PWM	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	3.6
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	0.0	7.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IL-13	2.8	0.0
B lymphocytes PWM	0.0	0.0	Lung fibroblast IFN gamma	6.0	17.3
B lymphocytes CD40L and IL-4	0.0	2.6	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	2.7	2.6	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	66.4	47.6	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	18.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells LPS	2.9	13.5	Dermal fibroblast IL-4	0.0	0.0
Dendritic cells anti-CD40	8.1	7.4	IBD Colitis 2	0.0	3.4
Monocytes rest	0.0	4.5	IBD Crohn's	0.0	2.3
Monocytes LPS	0.0	0.0	Colon	7.4	6.2
Macrophages rest	68.8	100.0	Lung	15.6	14.8
Macrophages LPS	31.6	63.3	Thymus	0.0	0.0
HUVEC none	0.0	0.0	Kidney	0.0	0.0
HUVEC starved	0.0	0.0			

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5267/Ag5268 Results from two experiments using different probe/primer sets show expression of this transcript in several normal and disease tissues; these results disagree with the data generated with the other two primer/probe sets. This observation suggests that the AG5267 and Ag5268 primer/probe sets may detect an isoform of the transcript with a wider expression pattern than that detected by Ag3228 and Ag3261. Please see Panel 4D for a discussion of the potential role of this protein in inflammation and its therapeutic utility. Ag3261 Significant expression of the NOV12a gene is limited to one psoriasis sample. Ag3228 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3261/Ag5267/Ag5268 Results from three experiments using this panel confirm the expression of this gene at low levels in the



brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders. Results from one  
5 experiment with the Ag3228 probe/primer set show low/undetectable levels (CTs>35) of expression in all the samples on this panel.

**General\_screening\_panel\_v1.4 Summary:** Ag3228 Highest expression of the NOV12a gene is seen in the cerebellum (CT = 30.4). Thus, expression of this gene can be used as a marker for cerebellum. Furthermore, this highly brain-preferential expression suggests a  
10 specific role for this gene product in the brain. The NOV12a gene encodes a protein with homology to neural cell adhesion molecules (NCAM). NCAM related proteins, such as Nr-CAM, play a critical role in neurite extension (ref. 1). Therefore, the introduction of ligands specific for this gene product, such as contactin, in directed brain regions may have utility in fostering focal neurite outgrowth and, thus may have utility in therapeutically countering  
15 neurite degeneration in neurodegenerative diseases such as Alzheimer's disease, ataxias, and Parkinson's disease.

Results from a second experiment with the Ag3261 probe and primer set are not included. The amp plot indicates that there were experimental difficulties with this run (Sakurai et al., J Cell Biol 154(6):1259-73, 2001).

**General\_screening\_panel\_v1.5 Summary:** Ag5268 Significant expression of the NOV12a gene is seen in the brain. Please see Panel 1.4 for discussion of utility of this gene in the brain. Significant expression is also seen in brain cancer cell lines. Thus, expression of this gene could be used to differentiate between brain derived samples and other samples on this  
20 panel.

**25** Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such  
30 as obesity and diabetes.

**Panel 2.2 Summary:** Ag3228 Significant expression of this gene is seen exclusively in an ovarian cancer sample (CT = 33.8). Therefore, expression of this gene may be used to distinguish ovarian cancers from the other samples on this panel. Furthermore, therapeutic

modulation of the activity of the protein encoded by this gene may be beneficial in the treatment of ovarian cancer.

**Panel 4.1D Summary:** Ag5267/Ag5268 Results from two experiments using different probe/primer sets are in good agreement with each other and show a similar overall pattern of expression as in Panel 4 but at much higher levels. This may result from differences in the two panels or from differences in the probe/primer sets used. The NOV12a transcript is expressed in LAK cells and treatment of the LAK cells with PMA and ionomycin upregulates the expression of this transcript. This transcript is also induced in activated EOL cells and in fibroblasts. The NOV12a gene encodes a putative NCAM, a type of cell surface protein often involved in cellular interaction, adhesion and signaling. Therefore, therapeutics designed with the protein encoded for this transcript could be important in the treatment of diseases such as asthma, emphysema, psoriasis and arthritis.

**Panel 4D Summary:** Ag3228/Ag3261 Results from two experiments using identical probe/primer sets are in good agreement. The NOV12a transcript is expressed in LAK cells and treatment of the LAK cells with PMA and ionomycin upregulates the expression of this transcript. The NOV12a gene encodes a putative NCAM, a type of cell surface protein often involved in cellular interaction, adhesion and signaling. Therefore, therapeutics designed with the protein encoded for this transcript could be important in the treatment of diseases such as asthma, emphysema, psoriasis and arthritis.

#### G. NOV13b and NOV13c: protein containing MAM and Ig domains

Expression of genes NOV13b and NOV13c was assessed using the primer-probe set Ag5267, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB, GC, GD and GE.

**Table GA.** Probe Name Ag5267

Primers	Sequences	Length	Start Position
Forward	5'-gcggtcccgaaca-3' (SEQ ID NO:224)	14	1166
Probe	TET-5'-cacgctggtctctcagtggca-3'-TAMRA (SEQ ID NO:225)	22	1197
Reverse	5'-gcctgctgccacacatt-3' (SEQ ID NO:226)	17	1227

**Table GB.** AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag5267, Run 230473002	Tissue Name	Rel. Exp.(%) Ag5267, Run 230473002
110967 COPD-F	36.9	112427 Match Control Psoriasis-F	84.1
110980 COPD-F	40.3	112418 Psoriasis-M	52.1
110968 COPD-M	36.1	112723 Match Control	10.4

		Psoriasis-M	
110977 COPD-M	0.0	112419 Psoriasis-M	61.1
110989 Emphysema-F	25.2	112424 Match Control Psoriasis-M	23.7
110992 Emphysema-F	12.2	112420 Psoriasis-M	100.0
110993 Emphysema-F	15.2	112425 Match Control Psoriasis-M	77.4
110994 Emphysema-F	11.1	104689 (MF) OA Bone-Backus	16.4
110995 Emphysema-F	18.9	104690 (MF) Adj "Normal" Bone-Backus	10.4
110996 Emphysema-F	0.0	104691 (MF) OA Synovium-Backus	9.6
110997 Asthma-M	21.2	104692 (BA) OA Cartilage-Backus	2.7
111001 Asthma-F	20.4	104694 (BA) OA Bone-Backus	21.2
111002 Asthma-F	22.4	104695 (BA) Adj "Normal" Bone-Backus	11.4
111003 Atopic Asthma-F	49.0	104696 (BA) OA Synovium-Backus	31.0
111004 Atopic Asthma-F	49.3	104700 (SS) OA Bone-Backus	28.5
111005 Atopic Asthma-F	24.3	104701 (SS) Adj "Normal" Bone-Backus	38.2
111006 Atopic Asthma-F	6.5	104702 (SS) OA Synovium-Backus	65.5
111417 Allergy-M	26.4	117093 OA Cartilage Rep7	34.6
112347 Allergy-M	1.5	112672 OA Bone5	61.1
112349 Normal Lung-F	2.9	112673 OA Synovium5	20.3
112357 Normal Lung-F	51.4	112674 OA Synovial Fluid cells5	33.7
112354 Normal Lung-M	39.8	117100 OA Cartilage Rep14	16.3
112374 Crohns-F	25.0	112756 OA Bone9	5.1
112389 Match Control Crohns-F	18.2	112757 OA Synovium9	2.2
112375 Crohns-F	24.0	112758 OA Synovial Fluid Cells9	10.3
112732 Match Control Crohns-F	22.7	117125 RA Cartilage Rep2	34.9
112725 Crohns-M	3.1	113492 Bone2 RA	27.2
112387 Match Control Crohns-M	20.9	113493 Synovium2 RA	20.0
112378 Crohns-M	2.9	113494 Syn Fluid Cells RA	35.6
112390 Match Control Crohns-M	45.1	113499 Cartilage4 RA	52.1
112726 Crohns-M	50.0	113500 Bone4 RA	43.8
112731 Match Control Crohns-M	31.0	113501 Synovium4 RA	33.9
112380 Ulcer Col-F	20.6	113502 Syn Fluid Cells4 RA	19.5
112734 Match Control Ulcer Col-F	37.4	113495 Cartilage3 RA	15.5

112384 Ulcer Col-F	41.5	113496 Bone3 RA	24.0
112737 Match Control Ulcer Col-F	27.5	113497 Synovium3 RA	5.4
112386 Ulcer Col-F	25.2	113498 Syn Fluid Cells3 RA	23.3
112738 Match Control Ulcer Col-F	3.0	117106 Normal Cartilage Rep20	22.7
112381 Ulcer Col-M	4.5	113663 Bone3 Normal	4.4
112735 Match Control Ulcer Col-M	16.4	113664 Synovium3 Normal	0.7
112382 Ulcer Col-M	18.7	113665 Syn Fluid Cells3 Normal	1.0
112394 Match Control Ulcer Col-M	6.4	117107 Normal Cartilage Rep22	8.7
112383 Ulcer Col-M	16.6	113667 Bone4 Normal	22.1
112736 Match Control Ulcer Col-M	14.7	113668 Synovium4 Normal	40.1
112423 Psoriasis-F	40.6	113669 Syn Fluid Cells4 Normal	25.0

Table GC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag5267, Run 230510331	Tissue Name	Rel. Exp.(%) Ag5267, Run 230510331
AD 1 Hippo	27.5	Control (Path) 3 Temporal Ctx	13.1
AD 2 Hippo	48.3	Control (Path) 4 Temporal Ctx	36.3
AD 3 Hippo	26.6	AD 1 Occipital Ctx	26.8
AD 4 Hippo	20.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	60.7	AD 3 Occipital Ctx	16.4
AD 6 Hippo	66.4	AD 4 Occipital Ctx	33.9
Control 2 Hippo	60.7	AD 5 Occipital Ctx	50.0
Control 4 Hippo	22.4	AD 6 Occipital Ctx	11.9
Control (Path) 3 Hippo	0.6	Control 1 Occipital Ctx	4.7
AD 1 Temporal Ctx	62.4	Control 2 Occipital Ctx	38.4
AD 2 Temporal Ctx	56.3	Control 3 Occipital Ctx	19.2
AD 3 Temporal Ctx	20.0	Control 4 Occipital Ctx	12.6
AD 4 Temporal Ctx	42.9	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	66.0	Control (Path) 2 Occipital Ctx	9.7
AD 5 Sup Temporal Ctx	61.1	Control (Path) 3 Occipital Ctx	7.6
AD 6 Inf Temporal Ctx	37.9	Control (Path) 4 Occipital Ctx	18.4
AD 6 Sup Temporal Ctx	57.8	Control 1 Parietal Ctx	13.3
Control 1 Temporal Ctx	13.6	Control 2 Parietal Ctx	60.3
Control 2 Temporal Ctx	36.6	Control 3 Parietal Ctx	15.2
Control 3 Temporal Ctx	15.9	Control (Path) 1	80.1

		Parietal Ctx	
Control 3 Temporal Ctx	20.7	Control (Path) 2 Parietal Ctx	37.6
Control (Path) 1 Temporal Ctx	81.2	Control (Path) 3 Parietal Ctx	8.0
Control (Path) 2 Temporal Ctx	30.4	Control (Path) 4 Parietal Ctx	44.1

Table GD. General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5267, Run 232936653	Rel. Exp.(%) Ag5267, Run 254397162	Tissue Name	Rel. Exp.(%) Ag5267, Run 232936653	Rel. Exp.(%) Ag5267, Run 254397162
Adipose	0.1	1.6	Renal ca. TK-10	0.3	3.0
Melanoma* Hs688(A).T	0.4	5.2	Bladder	0.1	0.6
Melanoma* Hs688(B).T	0.6	9.5	Gastric ca. (liver met.) NCI-N87	0.0	0.4
Melanoma* M14	0.0	0.1	Gastric ca. KATO III	0.0	0.0
Melanoma* LOXIMVI	0.0	0.7	Colon ca. SW-948	0.0	0.0
Melanoma* SK- MEL-5	0.6	8.5	Colon ca. SW480	0.2	3.1
Squamous cell carcinoma SCC- 4	0.0	0.0	Colon ca.* (SW480 met) SW620	0.0	0.6
Testis Pool	0.1	1.1	Colon ca. HT29	0.0	0.0
Prostate ca.* (bone met) PC-3	0.0	0.1	Colon ca. HCT-116	0.2	2.7
Prostate Pool	0.1	1.3	Colon ca. CaCo-2	0.0	0.1
Placenta	0.0	0.2	Colon cancer tissue	0.1	1.2
Uterus Pool	0.2	2.7	Colon ca. SW1116	0.0	0.2
Ovarian ca. OVCAR-3	0.0	0.4	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK- OV-3	0.5	6.7	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	0.1	1.3	Colon Pool	0.3	3.2
Ovarian ca. OVCAR-5	0.1	2.0	Small Intestine Pool	0.4	5.5
Ovarian ca. IGROV-1	0.0	0.4	Stomach Pool	0.2	1.7
Ovarian ca. OVCAR-8	0.1	0.6	Bone Marrow Pool	0.2	3.0
Ovary	0.2	2.9	Fetal Heart	0.1	1.1
Breast ca. MCF- 7	0.0	0.0	Heart Pool	100.0	2.9
Breast ca. MDA- MB-231	0.5	7.5	Lymph Node Pool	0.5	7.2
Breast ca. BT 549	0.9	12.1	Fetal Skeletal Muscle	0.3	3.6

Breast ca. T47D	0.2	2.5	Skeletal Muscle Pool	0.1	0.9
Breast ca. MDA-N	0.3	3.0	Spleen Pool	0.1	1.4
Breast Pool	0.5	5.2	Thymus Pool	0.2	2.7
Trachea	0.1	0.9	CNS cancer (glio/astro) U87-MG	0.1	1.7
Lung	0.2	1.7	CNS cancer (glio/astro) U-118-MG	0.6	9.5
Fetal Lung	0.3	4.4	CNS cancer (neuro;met) SK-N-AS	0.0	0.4
Lung ca. NCI-N417	0.0	0.6	CNS cancer (astro) SF-539	0.0	0.7
Lung ca. LX-1	0.3	4.4	CNS cancer (astro) SNB-75	2.6	37.9
Lung ca. NCI-H146	0.0	0.1	CNS cancer (glio) SNB-19	0.0	0.5
Lung ca. SHP-77	0.0	0.3	CNS cancer (glio) SF-295	0.3	2.9
Lung ca. A549	0.3	3.5	Brain (Amygdala) Pool	0.2	2.3
Lung ca. NCI-H526	0.0	0.2	Brain (cerebellum)	8.5	100.0
Lung ca. NCI-H23	0.1	0.3	Brain (fetal)	3.3	45.4
Lung ca. NCI-H460	0.1	0.2	Brain (Hippocampus) Pool	0.4	4.1
Lung ca. HOP-62	0.1	1.2	Cerebral Cortex Pool	0.3	3.1
Lung ca. NCI-H522	1.2	20.7	Brain (Substantia nigra) Pool	0.2	2.8
Liver	0.0	0.2	Brain (Thalamus) Pool	0.3	4.5
Fetal Liver	0.0	0.3	Brain (whole)	1.0	10.8
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	0.3	5.3
Kidney Pool	0.8	9.2	Adrenal Gland	0.2	3.5
Fetal Kidney	0.1	0.7	Pituitary gland Pool	0.1	1.5
Renal ca. 786-0	0.0	0.3	Salivary Gland	0.0	0.4
Renal ca. A498	0.5	6.3	Thyroid (female)	0.0	0.4
Renal ca. ACHN	0.4	5.6	Pancreatic ca. CAPAN2	0.0	0.0
Renal ca. UO-31	0.2	2.2	Pancreas Pool	0.2	3.0

Table GE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5267, Run 230510063	Tissue Name	Rel. Exp.(%) Ag5267, Run 230510063
Secondary Th1 act	0.0	HUVEC IL-1beta	1.4

Secondary Th2 act	4.3	HUVEC IFN gamma	4.8
Secondary Tr1 act	3.4	HUVEC TNF alpha + IFN gamma	3.8
Secondary Th1 rest	0.6	HUVEC TNF alpha + IL4	0.4
Secondary Th2 rest	2.1	HUVEC IL-11	0.4
Secondary Tr1 rest	0.8	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.4
Primary Th2 act	0.6	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.8	Microvascular Dermal EC TNFalpha + IL-1beta	0.2
Primary Th1 rest	0.2	Bronchial epithelium TNFalpha + IL1beta	0.7
Primary Th2 rest	0.3	Small airway epithelium none	0.2
Primary Tr1 rest	0.2	Small airway epithelium TNFalpha + IL-1beta	1.6
CD45RA CD4 lymphocyte act	1.6	Coronary artery SMC rest	0.9
CD45RO CD4 lymphocyte act	3.4	Coronary artery SMC TNFalpha + IL-1beta	1.9
CD8 lymphocyte act	0.4	Astrocytes rest	8.8
Secondary CD8 lymphocyte rest	0.1	Astrocytes TNFalpha + IL-1beta	11.0
Secondary CD8 lymphocyte act	0.2	KU-812 (Basophil) rest	1.4
CD4 lymphocyte none	0.9	KU-812 (Basophil) PMA/ionomycin	9.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.8	CCD1106 (Keratinocytes) none	4.8
LAK cells rest	6.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.2
LAK cells IL-2	3.3	Liver cirrhosis	2.0
LAK cells IL-2+IL-12	1.2	NCI-H292 none	0.4
LAK cells IL-2+IFN gamma	0.7	NCI-H292 IL-4	1.8
LAK cells IL-2+ IL-18	0.6	NCI-H292 IL-9	1.4
LAK cells PMA/ionomycin	41.8	NCI-H292 IL-13	2.8
NK Cells IL-2 rest	9.0	NCI-H292 IFN gamma	5.6
Two Way MLR 3 day	2.9	HPAEC none	0.2
Two Way MLR 5 day	3.5	HPAEC TNF alpha + IL-1 beta	4.9
Two Way MLR 7 day	2.7	Lung fibroblast none	46.7
PBMC rest	0.3	Lung fibroblast TNF alpha + IL-1 beta	9.7
PBMC PWM	0.0	Lung fibroblast IL-4	23.5
PBMC PHA-L	1.2	Lung fibroblast IL-9	32.5
Ramos (B cell) none	0.0	Lung fibroblast IL-13	18.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	73.7
B lymphocytes PWM	0.7	Dermal fibroblast CCD1070 rest	1.8
B lymphocytes CD40L and IL-4	7.6	Dermal fibroblast CCD1070 TNF alpha	0.3

EOL-1 dbcAMP	7.7	Dermal fibroblast CCD1070 IL-1 beta	2.0
EOL-1 dbcAMP PMA/ionomycin	100.0	Dermal fibroblast IFN gamma	11.7
Dendritic cells none	0.5	Dermal fibroblast IL-4	4.0
Dendritic cells LPS	0.4	Dermal Fibroblasts rest	10.1
Dendritic cells anti-CD40	0.4	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.7
Monocytes LPS	3.0	Colon	0.6
Macrophages rest	0.0	Lung	0.7
Macrophages LPS	0.7	Thymus	1.1
HUVEC none	0.7	Kidney	2.4
HUVEC starved	1.2		

**AI\_comprehensive\_panel\_v1.0 Summary:** Ag5267 This panel confirms the expression of the NOV13b gene in several normal and disease tissues with relevance to human immune function. Please see Panel 4.1D for a discussion of the potential role of this protein in inflammation and its therapeutic utility.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag5267 Results from this experiment show the expression of this gene at low levels in the brains of several individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment.

The NOV13b gene encodes a protein with homology to neural cell adhesion molecules (NCAM). NCAM related proteins, such as Nr-CAM, play a critical role in neurite extension (ref. 1). Therefore, the introduction of ligands specific for this gene product, such as contactin, in directed brain regions may have utility in fostering focal neurite outgrowth and, thus may have utility in therapeutically countering neurite degeneration in neurodegenerative diseases such as Alzheimer's disease, ataxias, and Parkinson's disease (Sakurai et al., 2001).

**General\_screening\_panel\_v1.5 Summary:** Ag5267 Results from one experiment are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panel 4.1D Summary:** Ag5267 The NOV13b transcript is expressed in LAK cells and treatment of the LAK cells with PMA and ionomycin upregulates the expression of this transcript. This transcript is also induced in activated EOL cells and in fibroblasts. The NOV13b gene encodes a putative NCAM, a type of cell surface protein often involved in cellular interaction, adhesion and signaling. Therefore, therapeutics designed with the protein encoded for this transcript could be important in the treatment of diseases such as asthma, emphysema, psoriasis and arthritis.



**H. NOV18a: Adipophilin**

Expression of gene NOV18a was assessed using the primer-probe set Ag5737, described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB and HC.

**Table HA.** Probe Name Ag5737

Primers	Sequences	Length	Start Position
Forward	5'-gagacagcagggctactttgt-3' (SEQ ID NO:227)	21	611
Probe	TET-5'-cacctggcctacgagcactctgtg-3'-TAMRA (SEQ ID NO:228)	24	663
Reverse	5'-gtgtttgctctgcctcagttt-3' (SEQ ID NO:229)	21	690

5

**Table HB.** General\_screening\_panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5737, Run 245385011	Tissue Name	Rel. Exp.(%) Ag5737, Run 245385011
Adipose	7.5	Renal ca. TK-10	0.4
Melanoma* Hs688(A).T	0.0	Bladder	57.8
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	3.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.7
Melanoma* SK-MEL-5	2.1	Colon ca. SW480	1.8
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.6	Colon ca. HT29	1.8
Prostate ca.* (bone met) PC-3	0.4	Colon ca. HCT-116	6.2
Prostate Pool	1.7	Colon ca. CaCo-2	1.4
Placenta	0.0	Colon cancer tissue	1.4
Uterus Pool	0.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	7.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.5
Ovarian ca. OVCAR-5	25.7	Small Intestine Pool	6.0
Ovarian ca. IGROV-1	2.3	Stomach Pool	3.7
Ovarian ca. OVCAR-8	8.2	Bone Marrow Pool	0.3
Ovary	5.2	Fetal Heart	11.8
Breast ca. MCF-7	7.9	Heart Pool	5.6
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	1.0
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	1.4
Breast ca. T47D	1.9	Skeletal Muscle Pool	100.0
Breast ca. MDA-N	0.0	Spleen Pool	1.7
Breast Pool	0.2	Thymus Pool	4.0
Trachea	12.6	CNS cancer (glio/astro) U87-MG	0.0
Lung	1.2	CNS cancer (glio/astro) U-118-MG	0.7

Fetal Lung	3.8	CNS cancer (neuro;met) SK-N-AS	0.7
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.4	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.4	CNS cancer (glio) SNB-19	4.8
Lung ca. SHP-77	1.6	CNS cancer (glio) SF-295	0.2
Lung ca. A549	0.0	Brain (Amygdala) Pool	2.0
Lung ca. NCI-H526	0.4	Brain (cerebellum)	0.9
Lung ca. NCI-H23	2.0	Brain (fetal)	0.7
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	1.6
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	1.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	2.8
Liver	6.3	Brain (Thalamus) Pool	4.2
Fetal Liver	11.0	Brain (whole)	2.2
Liver ca. HepG2	0.7	Spinal Cord Pool	6.0
Kidney Pool	5.8	Adrenal Gland	1.5
Fetal Kidney	5.6	Pituitary gland Pool	4.1
Renal ca. 786-0	0.2	Salivary Gland	12.7
Renal ca. A498	0.0	Thyroid (female)	2.0
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.2
Renal ca. UO-31	0.0	Pancreas Pool	10.7

Table HC. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag5737, Run 244646641	Tissue Name	Rel. Exp.(%) Ag5737, Run 244646641
97457_Patient-02go_adipose	24.1	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	14.7	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	93.3	94713_Donor 2 AD - B_adipose	2.4
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	6.8	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	5.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal	29.7	94735_Donor 3 AD - C_adipose	0.0

muscle			
97497_Patient-11ut_uterus	0.0	77138_Liver_HepG2untreated	6.3
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	34.4	81735_Small Intestine	16.7
97501_Patient-12sk_skeletal muscle	100.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	0.0
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

**General\_screening\_panel\_v1.5 Summary:** Ag5737 Expression of the NOV18a gene is highest in adult skeletal muscle (CT = 28.2) and is much lower in fetal skeletal muscle (CT = 34.4). Thus, expression of this gene may be used to distinguish adult and fetal skeletal muscle. Among other tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in adipose, liver, heart, pancreas, adrenal gland, pituitary gland and thyroid. The NOV18a gene encodes a protein with homology to adipophilin. Adipophilin is believed to be involved in fatty acid uptake in adipocytes and is associated with lipid globules in many types of animal cells (ref. 1-2). This gene product may be a critical player in lipid homeostasis; therefore, therapeutic modulation of the activity of the NOV18a gene or its protein product may be a treatment for metabolic disease, including obesity and diabetes.

In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Finally, expression of this gene appears to be primarily associated with normal tissues as compared to cancer cell lines. NOV18a gene expression appears to be down regulated in CNS and renal cancer cell lines. Therefore, use of the NOV18a gene product as a protein therapeutic may be of benefit in the treatment of CNS and renal cancers (Londos et al., Semin Cell Dev Biol. 10(1):51-8, 1999; Serrero et al., Biochim Biophys Acta. 1488(3):245-54, 2000).

**Panel 5 Islet Summary:** Ag5737 The NOV18a gene is moderately expressed in the pancreatic islets of Langerhans (CT = 32.4), as well as in a sample of skeletal muscle. The

- NOV18a gene encodes a protein with homology to adipophilin, which is believed to be involved in fatty acid uptake in adipocytes and is associated with lipid globules in many types of animal cells. Lipid homeostasis is critically involved in insulin secretion by islet beta cells. Therapeutic modulation of this gene product may be a treatment for the beta cell secretory defect in Type 2 diabetes (Unger and Orci, FASEB J. 15(2):312-21, 2001; Unger and Zhou, Diabetes. 50 Suppl 1:S118-21, 2001).

#### I. NOV19a: ##

- Expression of gene NOV19a was assessed using the primer-probe set Ag3549, described in Table IA.

Table IA. Probe Name Ag3549

Primers	Sequences	Length	Start Position
Forward	5'-ccccaggaagaggacata-3' (SEQ ID NO:230)	20	649
Probe	TET-5'-tgacacaggttctccctctgcaaaa-3'-TAMRA (SEQ ID NO:231)	25	669
Reverse	5'-ctgaggaggacctggacagt-3' (SEQ ID NO:232)	20	725

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3549 Expression of the NOV19a gene is low/undetectable in all samples on this panel (CTs>35).

- General\_screening\_panel\_v1.4 Summary:** Ag3549 Expression of the NOV19a gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 4D Summary:** Ag3549 Expression of the NOV19a gene is low/undetectable in all samples on this panel (CTs>35).

#### J. NOV20a: ##

- Expression of gene NOV20a was assessed using the primer-probe set Ag3866, described in Table JA.

Table JA. Probe Name Ag3866

Primers	Sequences	Length	Start Position
Forward	5'-gaactcctggcctcaagatc-3' (SEQ ID NO:233)	20	58
Probe	TET-5'-aaaggcccagcccctctcttttct-3'-TAMRA (SEQ ID NO:234)	24	96
Reverse	5'-aggaaggaaggaaggaagga-3' (SEQ ID NO:235)	20	116

- CNS\_neurodegeneration\_v1.0 Summary:** Ag3866 Expression of the CG59846-01 gene is low/undetectable in all samples on this panel (CTs>35). The amp plot indicates that there is a high probability of a probe failure.

**General\_screening\_panel\_v1.4 Summary:** Ag3866 Expression of the CG59846-01 gene is low/undetectable in all samples on this panel (CTs>35). The amp plot indicates that there is a high probability of a probe failure.

**Panel 2.2 Summary:** Ag3866 Expression of the CG59846-01 gene is low/undetectable in all samples on this panel (CTs>35). The amp plot indicates that there is a high probability of a probe failure.

**Panel 4.1D Summary:** Ag3866 Expression of the CG59846-01 gene is low/undetectable in all samples on this panel (CTs>35). The amp plot indicates that there is a high probability of a probe failure.

10

#### **K. NOV21a: Neurotransmission-associated protein**

Expression of gene NOV21a was assessed using the primer-probe set Ag675, described in Table KA.

Table KA. Probe Name Ag675

Primers	Sequences	Length	Start Position
Forward	5'-ccttagctaagcaggtcatgaa-3' (SEQ ID NO:236)	22	659
Probe	TET-5'-ctagtgccatccctgccaatctagt-3'-TAMRA (SEQ ID NO:237)	26	685
Reverse	5'-attgaaggaagcctcgatca-3' (SEQ ID NO:238)	20	731

**Panel 1.1 Summary:** Ag675 Expression of the NOV21a gene is low/undetectable in all samples on this panel (CTs>35).

#### **L. NOV21n: Neurotransmission-associated Protein (NTAP)**

Expression of gene NOV21n was assessed using the primer-probe set Ag675, described in Table LA.

20

Table LA. Probe Name Ag675

Primers	Sequences	Length	Start Position
Forward	5'-ccttagctaagcaggtcatgaa-3' (SEQ ID NO:239)	22	554
Probe	TET-5'-ctagtgccatccctgccaatctagt-3'-TAMRA (SEQ ID NO:240)	26	580
Reverse	5'-attgaaggaagcctcgatca-3' (SEQ ID NO:241)	20	626

**Panel 1.1 Summary:** Ag675 Expression of the NOV21a gene is low/undetectable in all samples on this panel (CTs>35).

**25 M. NOV22a: drebrin**

Expression of gene NOV22a was assessed using the primer-probe set Ag3946, described in Table MA.

**Table MA.** Probe Name Ag3946

Primers	Sequences	Length	Start Position
Forward	5'-ggtgattccacacatcctt-3' (SEQ ID NO:242)	20	1583
Probe	TET-5'-acctcccagacagcttggtctt-3'-TAMRA (SEQ ID NO:243)	24	1616
Reverse	5'-cagggcttggtcagatc-3' (SEQ ID NO:244)	19	1651

**Panel CNS\_1 Summary:** Ag3946 Expression of the NOV22a gene is

- 5 low/undetectable in all samples on this panel (CTs>35).

#### N. NOV23a: UNC5H2 homolog

- Expression of gene NOV23a was assessed using the primer-probe set Ag3546, described in Table NA. Results of the RTQ-PCR runs are shown in Tables NB, NC, ND, NE and NF.

**Table NA.** Probe Name Ag3546

Primers	Sequences	Length	Start Position
Forward	5'-ccatgaacagatcctccaagt-3' (SEQ ID NO:245)	21	2447
Probe	TET-5'-tgaacgagaaaccatcacttcttcg-3'-TAMRA (SEQ ID NO:246)	26	2489
Reverse	5'-ggaaagtgtgtcctcttg-3' (SEQ ID NO:247)	21	2515

**Table NB.** CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3546, Run 210629739	Tissue Name	Rel. Exp.(%) Ag3546, Run 210629739
AD 1 Hippo	8.0	Control (Path) 3 Temporal Ctx	6.6
AD 2 Hippo	28.1	Control (Path) 4 Temporal Ctx	33.4
AD 3 Hippo	3.7	AD 1 Occipital Ctx	15.1
AD 4 Hippo	7.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	85.9	AD 3 Occipital Ctx	7.6
AD 6 Hippo	18.7	AD 4 Occipital Ctx	37.6
Control 2 Hippo	27.4	AD 5 Occipital Ctx	47.6
Control 4 Hippo	6.5	AD 6 Occipital Ctx	21.5
Control (Path) 3 Hippo	4.2	Control 1 Occipital Ctx	4.2
AD 1 Temporal Ctx	10.9	Control 2 Occipital Ctx	59.9
AD 2 Temporal Ctx	75.8	Control 3 Occipital Ctx	25.0
AD 3 Temporal Ctx	6.3	Control 4 Occipital Ctx	4.7
AD 4 Temporal Ctx	26.1	Control (Path) 1 Occipital Ctx	98.6
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2	17.3

		Occipital Ctx	
AD 5 Sup Temporal Ctx	32.3	Control (Path) 3 Occipital Ctx	4.7
AD 6 Inf Temporal Ctx	38.2	Control (Path) 4 Occipital Ctx	22.4
AD 6 Sup Temporal Ctx	41.8	Control 1 Parietal Ctx	7.6
Control 1 Temporal Ctx	5.5	Control 2 Parietal Ctx	36.9
Control 2 Temporal Ctx	40.9	Control 3 Parietal Ctx	23.0
Control 3 Temporal Ctx	21.6	Control (Path) 1 Parietal Ctx	97.9
Control 3 Temporal Ctx	18.6	Control (Path) 2 Parietal Ctx	32.8
Control (Path) 1 Temporal Ctx	72.7	Control (Path) 3 Parietal Ctx	4.3
Control (Path) 2 Temporal Ctx	48.3	Control (Path) 4 Parietal Ctx	57.0

Table NC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3546, Run 213391156	Tissue Name	Rel. Exp.(%) Ag3546, Run 213391156
Adipose	0.3	Renal ca. TK-10	0.7
Melanoma* Hs688(A).T	0.0	Bladder	0.4
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.3
Melanoma* M14	3.4	Gastric ca. KATO III	0.4
Melanoma* LOXIMVI	0.6	Colon ca. SW-948	0.2
Melanoma* SK-MEL-5	0.4	Colon ca. SW480	1.9
Squamous cell carcinoma SCC-4	0.8	Colon ca.* (SW480 met) SW620	0.4
Testis Pool	0.8	Colon ca. HT29	0.5
Prostate ca.* (bone met) PC-3	0.4	Colon ca. HCT-116	1.3
Prostate Pool	15.8	Colon ca. CaCo-2	0.2
Placenta	0.0	Colon cancer tissue	0.7
Uterus Pool	0.5	Colon ca. SW1116	0.4
Ovarian ca. OVCAR-3	0.3	Colon ca. Colo-205	0.3
Ovarian ca. SK-OV-3	0.2	Colon ca. SW-48	0.6
Ovarian ca. OVCAR-4	0.5	Colon Pool	6.6
Ovarian ca. OVCAR-5	0.3	Small Intestine Pool	6.9
Ovarian ca. IGROV-1	16.3	Stomach Pool	3.0
Ovarian ca. OVCAR-8	17.7	Bone Marrow Pool	0.4
Ovary	0.1	Fetal Heart	0.4
Breast ca. MCF-7	0.2	Heart Pool	1.6
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	5.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.4
Breast ca. T47D	1.2	Skeletal Muscle Pool	0.2
Breast ca. MDA-N	3.0	Spleen Pool	0.4

Breast Pool	14.6	Thymus Pool	5.5
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.5
Lung	0.0	CNS cancer (glio/astro) U-118-MG	3.3
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.6
Lung ca. LX-1	0.4	CNS cancer (astro) SNB-75	4.7
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	14.2
Lung ca. SHP-77	8.1	CNS cancer (glio) SF-295	0.7
Lung ca. A549	9.4	Brain (Amygdala) Pool	15.7
Lung ca. NCI-H526	0.6	Brain (cerebellum)	12.2
Lung ca. NCI-H23	0.8	Brain (fetal)	100.0
Lung ca. NCI-H460	6.3	Brain (Hippocampus) Pool	28.7
Lung ca. HOP-62	0.9	Cerebral Cortex Pool	51.4
Lung ca. NCI-H522	0.9	Brain (Substantia nigra) Pool	37.4
Liver	0.0	Brain (Thalamus) Pool	35.1
Fetal Liver	0.8	Brain (whole)	47.3
Liver ca. HepG2	0.0	Spinal Cord Pool	4.7
Kidney Pool	4.9	Adrenal Gland	3.0
Fetal Kidney	28.7	Pituitary gland Pool	6.3
Renal ca. 786-0	2.0	Salivary Gland	0.1
Renal ca. A498	0.2	Thyroid (female)	1.4
Renal ca. ACHN	0.5	Pancreatic ca. CAPAN2	0.9
Renal ca. UO-31	2.0	Pancreas Pool	7.3

Table ND. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3546, Run 173762016	Tissue Name	Rel. Exp.(%) Ag3546, Run 173762016
Normal Colon	26.4	Kidney Margin (OD04348)	100.0
Colon cancer (OD06064)	0.0	Kidney malignant cancer (OD06204B)	0.0
Colon Margin (OD06064)	0.0	Kidney normal adjacent tissue (OD06204E)	13.0
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	58.2
Colon Margin (OD06159)	1.6	Kidney Margin (OD04450-03)	51.4
Colon cancer (OD06297-04)	3.8	Kidney Cancer 8120613	14.5
Colon Margin (OD06297-05)	7.2	Kidney Margin 8120614	22.5
CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer 9010320	0.0
CC Margin (ODO3921)	5.5	Kidney Margin 9010321	12.6
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	6.3
Lung Margin (OD06104)	11.5	Kidney Margin 8120608	12.0
Colon mets to lung	0.0	Normal Uterus	17.7



(OD04451-01)			
Lung Margin (OD04451-02)	7.3	Uterine Cancer 064011	1.8
Normal Prostate	11.1	Normal Thyroid	0.0
Prostate Cancer (OD04410)	16.5	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	32.5	Thyroid Cancer A302152	0.0
Normal Ovary	0.0	Thyroid Margin A302153	1.4
Ovarian cancer (OD06283-03)	3.6	Normal Breast	4.4
Ovarian Margin (OD06283-07)	0.0	Breast Cancer (OD04566)	0.0
Ovarian Cancer 064008	4.9	Breast Cancer 1024	0.0
Ovarian cancer (OD06145)	0.0	Breast Cancer (OD04590-01)	0.0
Ovarian Margin (OD06145)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Ovarian cancer (OD06455-03)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Ovarian Margin (OD06455-07)	0.0	Breast Cancer 064006	0.0
Normal Lung	0.0	Breast Cancer 9100266	1.7
Invasive poor diff. lung adeno (ODO4945-01)	0.0	Breast Margin 9100265	0.0
Lung Margin (ODO4945-03)	3.7	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Breast cancer (OD06083)	0.0
Lung Cancer (OD05014A)	0.0	Breast cancer node metastasis (OD06083)	0.0
Lung Margin (OD05014B)	7.4	Normal Liver	0.0
Lung cancer (OD06081)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD06081)	1.5	Liver Cancer 1025	0.0
Lung Cancer (OD04237-01)	0.0	Liver Cancer 6004-T	2.1
Lung Margin (OD04237-02)	10.8	Liver Tissue 6004-N	0.0
Ocular Melanoma Metastasis	0.0	Liver Cancer 6005-T	0.0
Ocular Melanoma Margin (Liver)	0.0	Liver Tissue 6005-N	0.0
Melanoma Metastasis	0.0	Liver Cancer 064003	0.0
Melanoma Margin (Lung)	0.0	Normal Bladder	0.0
Normal Kidney	21.8	Bladder Cancer 1023	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	40.3	Bladder Cancer A302173	0.0
Kidney Margin (OD04338)	3.2	Normal Stomach	9.9
Kidney Ca Nuclear grade 1/2 (OD04339)	94.6	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	19.9	Stomach Margin 9060396	0.0
Kidney Ca, Clear cell type (OD04340)	2.1	Gastric Cancer 9060395	0.0
Kidney Margin (OD04340)	26.1	Stomach Margin 9060394	0.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 064005	0.0

Table NE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3546, Run 166453846	Tissue Name	Rel. Exp.(%) Ag3546, Run 166453846
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.2
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	1.8
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	1.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 _anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	10.7
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL- 1 beta	0.2
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0

Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	1.9
Monocytes rest	0.0	IBD Crohn's	7.0
Monocytes LPS	0.0	Colon	100.0
Macrophages rest	0.0	Lung	0.2
Macrophages LPS	0.0	Thymus	30.6
HUVEC none	0.0	Kidney	6.3
HUVEC starved	0.0		

Table NF. Panel CNS\_1

Tissue Name	Rel. Exp.(%) Ag3546, Run 171647125	Tissue Name	Rel. Exp.(%) Ag3546, Run 171647125
BA4 Control	47.3	BA17 PSP	33.7
BA4 Control2	49.0	BA17 PSP2	13.0
BA4 Alzheimer's2	10.5	Sub Nigra Control	14.8
BA4 Parkinson's	58.6	Sub Nigra Control2	24.5
BA4 Parkinson's2	90.1	Sub Nigra Alzheimer's2	15.5
BA4 Huntington's	49.0	Sub Nigra Parkinson's2	47.6
BA4 Huntington's2	13.5	Sub Nigra Huntington's	31.4
BA4 PSP	16.5	Sub Nigra Huntington's2	23.5
BA4 PSP2	41.5	Sub Nigra PSP2	7.4
BA4 Depression	14.5	Sub Nigra Depression	0.9
BA4 Depression2	14.0	Sub Nigra Depression2	5.2
BA7 Control	62.9	Glob Palladus Control	7.0
BA7 Control2	29.5	Glob Palladus Control2	11.4
BA7 Alzheimer's2	13.7	Glob Palladus Alzheimer's	9.9
BA7 Parkinson's	21.0	Glob Palladus Alzheimer's2	3.5
BA7 Parkinson's2	61.1	Glob Palladus Parkinson's	66.9
BA7 Huntington's	63.7	Glob Palladus Parkinson's2	6.6
BA7 Huntington's2	73.7	Glob Palladus PSP	3.3
BA7 PSP	70.2	Glob Palladus PSP2	8.3
BA7 PSP2	48.0	Glob Palladus	3.6

		Depression	
BA7 Depression	17.8	Temp Pole Control	22.5
BA9 Control	34.6	Temp Pole Control2	83.5
BA9 Control2	100.0	Temp Pole Alzheimer's	15.2
BA9 Alzheimer's	7.7	Temp Pole Alzheimer's2	10.7
BA9 Alzheimer's2	30.1	Temp Pole Parkinson's	37.9
BA9 Parkinson's	51.4	Temp Pole Parkinson's2	35.8
BA9 Parkinson's2	67.8	Temp Pole Huntington's	59.0
BA9 Huntington's	72.2	Temp Pole PSP	9.4
BA9 Huntington's2	24.8	Temp Pole PSP2	10.4
BA9 PSP	23.8	Temp Pole Depression2	11.1
BA9 PSP2	7.4	Cing Gyr Control	73.2
BA9 Depression	11.7	Cing Gyr Control2	33.7
BA9 Depression2	18.4	Cing Gyr Alzheimer's	25.0
BA17 Control	65.1	Cing Gyr Alzheimer's2	14.5
BA17 Control2	75.8	Cing Gyr Parkinson's	30.4
BA17 Alzheimer's2	19.8	Cing Gyr Parkinson's2	35.4
BA17 Parkinson's	51.4	Cing Gyr Huntington's	72.2
BA17 Parkinson's2	67.8	Cing Gyr Huntington's2	21.2
BA17 Huntington's	48.6	Cing Gyr PSP	13.8
BA17 Huntington's2	37.1	Cing Gyr PSP2	7.5
BA17 Depression	15.4	Cing Gyr Depression	6.9
BA17 Depression2	35.4	Cing Gyr Depression2	14.2

- CNS\_neurodegeneration\_v1.0 Summary:** Ag3546 This panel confirms the expression of the NOV23a gene at significant levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between
- 5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

- General\_screening\_panel\_v1.4 Summary:** Ag3546 Highest expression of the NOV23a gene is detected in fetal brain (CT=25.2). In addition, this gene is expressed at high
- 10 levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord (CTs=26-29). Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

The NOV23a gene encodes a homologue of rat UNC5H2 gene. Members of UNC5H are membrane receptors for netrin-1 and crucial for axon guidance and neuronal migration. Netrins are a family of chemotropic factors that guide axon outgrowth during development. In situ hybridization has revealed that the netrin 1 receptors, DCC1 and UNC5H2 mRNAs are expressed by normal adult retinal ganglion cells (RGCs). In addition, expression of DCC1 and UNC5H2 mRNA is down regulated in RGCs that has undergone axotomy. Thus, netrin-1, DCC, and UNC5H2 may contribute to regulating the regenerative capacity of adult RGCs (Ref.1). Thus, high expression of the NOV23a gene in both fetal and adult brain, suggests this gene product may also play a role in the regenerative capacity of adult RGCs.

Recently, it was shown that netrin-1 receptors UNC5H (UNC5H1, UNC5H2, UNC5H3) also act as dependence receptors. They induce apoptosis, but this effect is blocked in the presence of netrin-1. Thus, during development of the nervous system, the presence of netrin-1 is crucial to maintain survival of UNC5H- and DCC-expressing neurons, especially in the ventricular zone of the brainstem (Ref. 2). Therefore, the NOV23a gene product along with Netrin 1 may be important in the survival of the neurons.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Significant expression is also detected in fetal liver and fetal lung. Interestingly, this gene is expressed at much higher levels in fetal (CTs = 32-34.9) when compared to adult liver and lung (CTs = 40). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver or lung, respectively (Ellezam et al., Exp Neurol 168(1):105-15, 2001; Llambi et al., EMBO J 20(11):2715-22, 2001).

**Panel 2.2 Summary:** Ag3546 Expression of the NOV23a gene on this panel is seen primarily in kidney derived tissue (CTs=32-33). Thus, expression of this gene could be used to differentiate between kidney derived samples and other samples on this panel.

**Panel 4D Summary:** Ag3546 Expression of the NOV23a gene is highest in normal colon (CT=28.3). Therefore, expression of this gene may be used to distinguish colon from the other tissues on this panel. Furthermore, expression of this gene is decreased in colon samples from patients with IBD colitis and Crohn's disease relative to normal colon. Therefore, therapeutic modulation of the activity of the protein encoded by this gene may be useful in the treatment of inflammatory bowel disease.

**Panel CNS\_1 Summary:** Ag3546 This panel confirms the expression of the NOV23a gene at significant levels in the brains of an independent group of individuals. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

5

#### O. NOV24a: Trypsin inhibitor

Expression of gene NOV24a was assessed using the primer-probe set Ag3485, described in Table OA. Results of the RTQ-PCR runs are shown in Tables OB and OC.

Table OA. Probe Name Ag3485

Primers	Sequences	Length	Start Position
Forward	5'-gccttcacagctgatgagatac-3' (SEQ ID NO:248)	22	349
Probe	TET-5'-aacctctccatccattctggccagta-3'-TAMRA (SEQ ID NO:249)	26	380
Reverse	5'-tcagaccaggacttcatgagat-3' (SEQ ID NO:250)	22	420

10

Table OB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3485, Run 217215038	Tissue Name	Rel. Exp.(%) Ag3485, Run 217215038
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.4
Melanoma* M14	0.5	Gastric ca. KATO III	0.0
Melanoma* LOXIMV1	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.9
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	3.1
Testis Pool	0.8	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	1.1	Colon cancer tissue	1.7
Uterus Pool	0.0	Colon ca. SW1116	1.1
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	2.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.8	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0

Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	4.1	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	7.9	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	4.6	CNS cancer (glio) SF-295	0.0
Lung ca. A549	100.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	7.9	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	1.1	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	1.3
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table OC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3485, Run 166441741	Tissue Name	Rel. Exp.(%) Ag3485, Run 166441741
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium	0.0

		TNFalpha + IL-1beta	
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	9.2
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	100.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		



**CNS\_neurodegeneration\_v1.0 Summary:** Ag3485 Expression of the NOV24a gene is low/undetectable in all samples on this panel (CTs>35).

**General\_screening\_panel\_v1.4 Summary:** Ag3485 Expression of the NOV24a gene is restricted to two samples derived from lung cancer cell lines (CTs=30-34). Thus, expression of this gene could be used to differentiate between these sample and other samples on this panel and as a marker to detect the presence of lung cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung cancer.

**Panel 4D Summary:** Ag3485 Expression of the NOV24a gene is restricted to normal lung tissue. This specific expression in lung derived tissue in both this panel and panel 1.4 suggests a role for this gene in the normal homeostasis of this tissue. Therapeutic modulation of the expression or function of this gene may be useful in maintaining or restoring normal function to the lung during inflammation.

#### 15 P. NOV26a and NOV26b: Ovostatin

Expression of gene NOV26a and variant NOV26b was assessed using the primer-probe set Ag1282, described in Table PA. Results of the RTQ-PCR runs are shown in Tables PB, PC, PD, PE and PF.

Table PA. Probe Name Ag1282

Primers	Sequences	Length	Start Position
Forward	5'-ttcgcaataaatccagtatggt-3' (SEQ ID NO:251)	22	3929
Probe	TET-5'-tgctatcaggatttactccaaccatgtca-3'-TAMRA (SEQ ID NO:252)	29	3968
Reverse	5'-ttggttttcaagctcttcaatgg-3' (SEQ ID NO:253)	22	3998

20

Table PB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag1282, Run 216588406	Tissue Name	Rel. Exp.(%) Ag1282, Run 216588406
Adipose	0.5	Renal ca. TK-10	1.1
Melanoma* Hs688(A).T	0.0	Bladder	0.4
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	1.7
Melanoma* M14	100.0	Gastric ca. KATO III	1.2
Melanoma* LOXIMVI	0.9	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	11.2	Colon ca. SW480	3.3
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	1.5
Testis Pool	5.3	Colon ca. HT29	0.4
Prostate ca.* (bone met)	0.0	Colon ca. HCT-116	2.5

PC-3			
Prostate Pool	0.3	Colon ca. CaCo-2	0.6
Placenta	0.4	Colon cancer tissue	0.5
Uterus Pool	0.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	2.1	Colon ca. Colo-205	0.1
Ovarian ca. SK-OV-3	0.7	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.2	Colon Pool	0.5
Ovarian ca. OVCAR-5	0.8	Small Intestine Pool	1.4
Ovarian ca. IGROV-1	0.2	Stomach Pool	0.2
Ovarian ca. OVCAR-8	0.1	Bone Marrow Pool	0.2
Ovary	0.6	Fetal Heart	1.6
Breast ca. MCF-7	0.3	Heart Pool	0.2
Breast ca. MDA-MB-231	0.6	Lymph Node Pool	0.4
Breast ca. BT 549	0.9	Fetal Skeletal Muscle	1.6
Breast ca. T47D	0.8	Skeletal Muscle Pool	0.1
Breast ca. MDA-N	45.1	Spleen Pool	0.8
Breast Pool	0.7	Thymus Pool	2.5
Trachea	0.5	CNS cancer (glio/astro) U87-MG	0.4
Lung	0.3	CNS cancer (glio/astro) U-118-MG	0.3
Fetal Lung	4.9	CNS cancer (neuro;met) SK-N-AS	0.5
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.1
Lung ca. LX-1	1.6	CNS cancer (astro) SNB-75	9.0
Lung ca. NCI-H146	2.3	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	1.1	CNS cancer (glio) SF-295	0.7
Lung ca. A549	1.1	Brain (Amygdala) Pool	0.9
Lung ca. NCI-H526	0.4	Brain (cerebellum)	0.6
Lung ca. NCI-H23	2.8	Brain (fetal)	2.9
Lung ca. NCI-H460	0.1	Brain (Hippocampus) Pool	1.0
Lung ca. HOP-62	0.7	Cerebral Cortex Pool	1.1
Lung ca. NCI-H522	2.0	Brain (Substantia nigra) Pool	0.8
Liver	0.0	Brain (Thalamus) Pool	1.1
Fetal Liver	0.9	Brain (whole)	0.8
Liver ca. HepG2	0.3	Spinal Cord Pool	0.5
Kidney Pool	1.0	Adrenal Gland	0.1
Fetal Kidney	3.6	Pituitary gland Pool	0.1
Renal ca. 786-0	0.6	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.1
Renal ca. ACHN	0.6	Pancreatic ca. CAPAN2	1.3
Renal ca. UO-31	0.1	Pancreas Pool	0.9

Table PC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1282, Run 167614616	Tissue Name	Rel. Exp.(%) Ag1282, Run 167614616
-------------	------------------------------------	-------------	------------------------------------

Liver adenocarcinoma	9.2	Kidney (fetal)	4.9
Pancreas	0.5	Renal ca. 786-0	0.3
Pancreatic ca. CAPAN 2	1.3	Renal ca. A498	1.9
Adrenal gland	0.3	Renal ca. RXF 393	0.1
Thyroid	0.3	Renal ca. ACHN	0.8
Salivary gland	0.1	Renal ca. UO-31	0.4
Pituitary gland	0.3	Renal ca. TK-10	2.1
Brain (fetal)	11.1	Liver	0.2
Brain (whole)	0.7	Liver (fetal)	1.4
Brain (amygdala)	1.1	Liver ca. (hepatoblast) HepG2	0.7
Brain (cerebellum)	0.3	Lung	0.5
Brain (hippocampus)	1.5	Lung (fetal)	8.0
Brain (substantia nigra)	2.3	Lung ca. (small cell) LX-1	1.3
Brain (thalamus)	2.4	Lung ca. (small cell) NCI-H69	13.2
Cerebral Cortex	1.0	Lung ca. (s.cell var.) SHP-77	5.9
Spinal cord	0.7	Lung ca. (large cell) NCI-H460	0.1
glio/astro U87-MG	0.4	Lung ca. (non-sm. cell) A549	0.3
glio/astro U-118-MG	0.4	Lung ca. (non-s.cell) NCI-H23	2.4
astrocytoma SW1783	0.4	Lung ca. (non-s.cell) HOP-62	1.4
neuro*; met SK-N-AS	0.7	Lung ca. (non-s.cl) NCI-H522	2.6
astrocytoma SF-539	0.3	Lung ca. (squamous) SW 900	2.5
astrocytoma SNB-75	7.8	Lung ca. (squamous) NCI-H596	49.7
glioma SNB-19	0.2	Mammary gland	2.5
glioma U251	1.2	Breast ca.* (pl.ef) MCF-7	0.5
glioma SF-295	0.6	Breast ca.* (pl.ef) MDA-MB-231	0.8
Heart (fetal)	5.1	Breast ca.* (pl.ef) T47D	0.3
Heart	0.7	Breast ca. BT-549	0.1
Skeletal muscle (fetal)	3.3	Breast ca. MDA-N	100.0
Skeletal muscle	0.5	Ovary	1.3
Bone marrow	3.9	Ovarian ca. OVCAR-3	2.6
Thymus	15.9	Ovarian ca. OVCAR-4	0.1
Spleen	1.2	Ovarian ca. OVCAR-5	2.7
Lymph node	4.6	Ovarian ca. OVCAR-8	0.1
Colorectal	2.5	Ovarian ca. IGROV-1	0.6
Stomach	0.6	Ovarian ca.* (ascites) SK-OV-3	1.3
Small intestine	1.8	Uterus	1.4
Colon ca. SW480	2.8	Placenta	0.2

Colon ca. * SW620(SW480 met)	5.3	Prostate	0.5
Colon ca. HT29	0.3	Prostate ca. * (bone met)PC-3	0.0
Colon ca. HCT-116	1.6	Testis	17.4
Colon ca. CaCo-2	0.8	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.2	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	3.1	Melanoma UACC-62	2.8
Gastric ca. * (liver met) NCI-N87	1.9	Melanoma M14	79.6
Bladder	0.9	Melanoma LOX IMVI	2.1
Trachea	0.3	Melanoma* (met) SK-MEL-5	9.9
Kidney	0.4	Adipose	1.9

Table PD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1282, Run 170849610	Tissue Name	Rel. Exp.(%) Ag1282, Run 170849610
Normal Colon	3.3	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	0.4	Kidney Cancer 8120613	0.1
CC Margin (ODO3866)	0.4	Kidney Margin 8120614	0.1
CC Gr.2 rectosigmoid (ODO3868)	1.3	Kidney Cancer 9010320	0.1
CC Margin (ODO3868)	0.1	Kidney Margin 9010321	0.1
CC Mod Diff (ODO3920)	3.6	Normal Uterus	0.3
CC Margin (ODO3920)	0.8	Uterus Cancer 064011	2.5
CC Gr.2 ascend colon (ODO3921)	0.9	Normal Thyroid	0.4
CC Margin (ODO3921)	0.4	Thyroid Cancer 064010	0.3
CC from Partial Hepatectomy (ODO4309) Mets	0.9	Thyroid Cancer A302152	0.2
Liver Margin (ODO4309)	0.2	Thyroid Margin A302153	0.4
Colon mets to lung (OD04451-01)	0.2	Normal Breast	1.8
Lung Margin (OD04451-02)	0.2	Breast Cancer (OD04566)	1.6
Normal Prostate 6546-1	1.8	Breast Cancer (OD04590-01)	0.7
Prostate Cancer (OD04410)	0.8	Breast Cancer Mets (OD04590-03)	2.3
Prostate Margin (OD04410)	1.1	Breast Cancer Metastasis (OD04655-05)	0.7
Prostate Cancer (OD04720-01)	0.7	Breast Cancer 064006	0.8
Prostate Margin (OD04720-02)	1.6	Breast Cancer 1024	1.2
Normal Lung 061010	2.6	Breast Cancer 9100266	0.9
Lung Met to Muscle (ODO4286)	0.3	Breast Margin 9100265	0.5
Muscle Margin (ODO4286)	0.2	Breast Cancer A209073	1.6
Lung Malignant Cancer	0.9	Breast Margin A209073	1.5

(OD03126)			
Lung Margin (OD03126)	0.7	Normal Liver	0.1
Lung Cancer (OD04404)	0.5	Liver Cancer 064003	0.3
Lung Margin (OD04404)	0.6	Liver Cancer 1025	0.1
Lung Cancer (OD04565)	0.8	Liver Cancer 1026	0.1
Lung Margin (OD04565)	0.4	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	25.5	Liver Tissue 6004-N	0.9
Lung Margin (OD04237-02)	0.7	Liver Cancer 6005-T	0.3
Ocular Mel Met to Liver (ODO4310)	100.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.4	Normal Bladder	1.1
Melanoma Mets to Lung (OD04321)	21.8	Bladder Cancer 1023	0.2
Lung Margin (OD04321)	0.5	Bladder Cancer A302173	6.5
Normal Kidney	0.5	Bladder Cancer (OD04718-01)	1.4
Kidney Ca, Nuclear grade 2 (OD04338)	0.4	Bladder Normal Adjacent (OD04718-03)	0.7
Kidney Margin (OD04338)	0.7	Normal Ovary	0.8
Kidney Ca Nuclear grade 1/2 (OD04339)	0.2	Ovarian Cancer 064008	2.9
Kidney Margin (OD04339)	0.1	Ovarian Cancer (OD04768-07)	3.8
Kidney Ca, Clear cell type (OD04340)	0.8	Ovary Margin (OD04768- 08)	0.2
Kidney Margin (OD04340)	0.6	Normal Stomach	1.2
Kidney Ca, Nuclear grade 3 (OD04348)	0.5	Gastric Cancer 9060358	0.8
Kidney Margin (OD04348)	0.2	Stomach Margin 9060359	0.7
Kidney Cancer (OD04622-01)	0.3	Gastric Cancer 9060395	0.8
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.2
Kidney Cancer (OD04450-01)	0.1	Gastric Cancer 9060397	0.5
Kidney Margin (OD04450-03)	0.1	Stomach Margin 9060396	0.2
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	2.5

Table PE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag1282, Run 169828985	Tissue Name	Rel. Exp.(%) Ag1282, Run 169828985
Secondary Th1 act	12.0	HUVEC IL-1beta	2.6
Secondary Th2 act	19.3	HUVEC IFN gamma	2.6
Secondary Tr1 act	21.6	HUVEC TNF alpha + IFN gamma	1.4
Secondary Th1 rest	7.5	HUVEC TNF alpha + IL4	3.1
Secondary Th2 rest	12.1	HUVEC IL-11	3.4
Secondary Tr1 rest	9.6	Lung Microvascular EC none	6.4
Primary Th1 act	20.0	Lung Microvascular EC TNFalpha + IL-1beta	9.5
Primary Th2 act	19.6	Microvascular Dermal EC none	10.4
Primary Tr1 act	24.3	Microvascular Dermal EC	10.2

		TNFalpha + IL-1beta	
Primary Th1 rest	23.8	Bronchial epithelium TNFalpha + IL1beta	3.4
Primary Th2 rest	16.3	Small airway epithelium none	0.0
Primary Tr1 rest	54.0	Small airway epithelium TNFalpha + IL-1beta	2.7
CD45RA CD4 lymphocyte act	7.6	Coronary artery SMC rest	0.3
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.1
CD8 lymphocyte act	23.3	Astrocytes rest	1.1
Secondary CD8 lymphocyte rest	9.3	Astrocytes TNFalpha + IL-1beta	2.4
Secondary CD8 lymphocyte act	16.4	KU-812 (Basophil) rest	14.5
CD4 lymphocyte none	3.6	KU-812 (Basophil) PMA/ionomycin	9.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	23.7	CCD1106 (Keratinocytes) none	4.6
LAK cells rest	7.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.4
LAK cells IL-2	26.8	Liver cirrhosis	2.2
LAK cells IL-2+IL-12	5.5	NCI-H292 none	10.6
LAK cells IL-2+IFN gamma	10.8	NCI-H292 IL-4	18.2
LAK cells IL-2+ IL-18	14.0	NCI-H292 IL-9	24.5
LAK cells PMA/ionomycin	1.5	NCI-H292 IL-13	14.8
NK Cells IL-2 rest	29.1	NCI-H292 IFN gamma	11.1
Two Way MLR 3 day	8.0	HPAEC none	3.3
Two Way MLR 5 day	7.0	HPAEC TNF alpha + IL-1 beta	5.3
Two Way MLR 7 day	9.6	Lung fibroblast none	1.1
PBMC rest	0.3	Lung fibroblast TNF alpha + IL-1 beta	0.6
PBMC PWM	4.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	12.7	Lung fibroblast IL-9	0.3
Ramos (B cell) none	8.2	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	11.5	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	10.7	Dermal fibroblast CCD1070 rest	5.4
B lymphocytes CD40L and IL-4	35.6	Dermal fibroblast CCD1070 TNF alpha	24.3
EOL-1 dbcAMP	9.2	Dermal fibroblast CCD1070 IL-1 beta	5.0
EOL-1 dbcAMP PMA/ionomycin	10.3	Dermal fibroblast IFN gamma	0.3
Dendritic cells none	1.2	Dermal fibroblast IL-4	1.6
Dendritic cells LPS	0.7	Dermal Fibroblasts rest	0.8
Dendritic cells anti-CD40	0.3	Neutrophils TNFa+LPS	0.3
Monocytes rest	1.1	Neutrophils rest	0.6
Monocytes LPS	2.4	Colon	8.4
Macrophages rest	2.2	Lung	5.1

Macrophages LPS	0.1	Thymus	100.0
HUVEC none	1.1	Kidney	1.3
HUVEC starved	3.6		

Table PF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1282, Run 166374199	Tissue Name	Rel. Exp.(%) Ag1282, Run 166374199
Secondary Th1 act	6.7	HUVEC IL-1beta	1.8
Secondary Th2 act	9.0	HUVEC IFN gamma	2.2
Secondary Tr1 act	15.8	HUVEC TNF alpha + IFN gamma	1.4
Secondary Th1 rest	5.0	HUVEC TNF alpha + IL4	1.1
Secondary Th2 rest	5.3	HUVEC IL-11	1.8
Secondary Tr1 rest	4.9	Lung Microvascular EC none	3.2
Primary Th1 act	18.6	Lung Microvascular EC TNFalpha + IL-1beta	7.2
Primary Th2 act	15.6	Microvascular Dermal EC none	12.7
Primary Tr1 act	16.2	Microvascular Dermal EC TNFalpha + IL-1beta	7.9
Primary Th1 rest	39.2	Bronchial epithelium TNFalpha + IL1beta	3.5
Primary Th2 rest	23.7	Small airway epithelium none	0.0
Primary Tr1 rest	33.0	Small airway epithelium TNFalpha + IL-1beta	4.8
CD45RA CD4 lymphocyte act	4.7	Coronary artery SMC rest	0.1
CD45RO CD4 lymphocyte act	9.4	Coronary artery SMC TNFalpha + IL-1beta	0.4
CD8 lymphocyte act	11.3	Astrocytes rest	0.3
Secondary CD8 lymphocyte rest	3.9	Astrocytes TNFalpha + IL-1beta	5.0
Secondary CD8 lymphocyte act	10.4	KU-812 (Basophil) rest	9.2
CD4 lymphocyte none	2.3	KU-812 (Basophil) PMA/ionomycin	9.2
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.9	CCD1106 (Keratinocytes) none	2.3
LAK cells rest	4.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.4
LAK cells IL-2	16.5	Liver cirrhosis	1.1
LAK cells IL-2+IL-12	6.7	Lupus kidney	0.7
LAK cells IL-2+IFN gamma	7.9	NCI-H292 none	15.3
LAK cells IL-2+ IL-18	6.8	NCI-H292 IL-4	23.3
LAK cells PMA/ionomycin	2.0	NCI-H292 IL-9	19.1
NK Cells IL-2 rest	18.6	NCI-H292 IL-13	15.3
Two Way MLR 3 day	5.4	NCI-H292 IFN gamma	13.7
Two Way MLR 5 day	3.6	HPAEC none	2.6
Two Way MLR 7 day	4.1	HPAEC TNF alpha + IL-1 beta	1.6

PBMC rest	1.0	Lung fibroblast none	0.0
PBMC PWM	21.9	Lung fibroblast TNF alpha + IL-1 beta	0.3
PBMC PHA-L	16.7	Lung fibroblast IL-4	0.1
Ramos (B cell) none	7.1	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	35.1	Lung fibroblast IL-13	0.0
B lymphocytes PWM	32.5	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	39.8	Dermal fibroblast CCD1070 rest	5.7
EOL-1 dbcAMP	6.4	Dermal fibroblast CCD1070 TNF alpha	37.1
EOL-1 dbcAMP PMA/ionomycin	4.5	Dermal fibroblast CCD1070 IL-1 beta	3.6
Dendritic cells none	0.7	Dermal fibroblast IFN gamma	1.2
Dendritic cells LPS	0.8	Dermal fibroblast IL-4	0.8
Dendritic cells anti-CD40	0.3	IBD Colitis 2	4.0
Monocytes rest	0.0	IBD Crohn's	2.0
Monocytes LPS	0.8	Colon	13.4
Macrophages rest	1.7	Lung	6.9
Macrophages LPS	0.0	Thymus	2.4
HUVEC none	2.7	Kidney	100.0
HUVEC starved	3.6		

**General\_screening\_panel\_v1.4 Summary:** Ag1282 Highest expression of the NOV26a gene is seen in a melanoma cell line. In addition, significantly higher levels of expression are seen in a breast cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of melanoma and breast cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of melanoma and breast cancers.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

In addition, this gene is expressed at much higher levels in fetal lung, liver and skeletal muscle tissue (CTs=27-29) when compared to expression in the adult counterpart (CTs=30-32). Thus, expression of this gene may be used to differentiate between the fetal and adult source of these tissue.



This molecule is a novel ovostatin that is also expressed at moderate in the regions of the CNS examined and may therefore be a target for the treatment of neurologic diseases.

**Panel 1.3D Summary:** Ag1282 Expression of the NOV26a gene is consistent with expression in Panel 1.4. The expression of this gene appears to be highest in a sample derived from a breast cancer cell line (MDA-N) (CT=26.9). In addition, there appears to be substantial expression in other samples derived from lung cancer cell lines and melanoma cell lines. Thus, the expression of this gene could be used to distinguish MDA-N cells from other samples in the panel. This gene encodes a novel ovostatin. Ovostatins are protease inhibitors that have been shown to support the growth of tumor cells in the absence of serum. They have also been shown to mediate accelerated fibroblast growth, collagen deposition and capillary formation. These activities suggest a role for this ovostatin homolog in tumor progression and proliferation. Thus, therapeutic targeting of this gene product may block the uncontrolled growth of cancer cells related to the action of the NOV26a gene. This could occur in any possible combination of cell growth, collagen deposition or capillary formation, especially in those cancer types like lung, breast and melanoma tumors where the gene is overexpressed in the tumor compared to the normal adjacent tissue. Please see Panel 1.4 for additional utility of this gene.

**Panel 2D Summary:** Ag1282 Highest expression of the NOV26a gene is seen in a sample derived from an ocular melanoma metastasis to the liver (CT=27). In addition, there appears to be substantial expression in other samples derived from lung cancers. Thus, expression of this gene could be used to distinguish liver cancer cells from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies could be of benefit in the treatment of liver or lung cancer.

**Panel 3D Summary:** Ag1282 Results from one experiment with the NOV26a gene are not included. The amp plot indicates that there were experimental difficulties with this run.

**Panels 4 and 4.1D Summary:** Ag1282 The NOV26a gene, an ovostatin-like protein, is related to ovostatin, a known inhibitor of proteinases of all four mechanistic classes, (serine proteinases, cysteine proteinases, aspartyl proteinases, and metalloproteinases) (see references). Highest expression of the gene is seen in the thymus and kidney (CTs=28-29). In addition, moderate to low levels of expression are seen in most of the samples on this panel. Thus, the NOV26a protein product may be useful as a therapeutic protein for the reduction of various proteolytic activities involved in inflammatory and autoimmune diseases such as, but not limited to, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic

obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis, wound healing, and infection (Saxena and Tayyab, Cell Mol Life Sci 53(1):13-23, 1997; Ofuji et al., Periodontal Clin Investig 14(2):13-22, 1992)

#### 5 Q. NOV28a: Laminin-type EGF like protein

Expression of gene NOV28a was assessed using the primer-probe set Ag399, described in Table QA. Results of the RTQ-PCR runs are shown in Tables QB, QC, QD, QE and QF.

Table QA. Probe Name Ag399

Primers	Sequences	Length	Start Position
Forward	5'-gcggccatgactgggtact-3' (SEQ ID NO:254)	19	1217
Probe	TET-5'-agcacacgggtcactgcgctctga-3'-TAMRA (SEQ ID NO:255)	23	1241
Reverse	5'-gcgattatctgcccttgatga-3' (SEQ ID NO:256)	21	1272

10

Table QB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag399, Run 225436712	Tissue Name	Rel. Exp.(%) Ag399, Run 225436712
AD 1 Hippo	17.4	Control (Path) 3 Temporal Ctx	10.8
AD 2 Hippo	33.7	Control (Path) 4 Temporal Ctx	35.8
AD 3 Hippo	18.4	AD 1 Occipital Ctx	33.9
AD 4 Hippo	14.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	16.0
AD 6 Hippo	52.9	AD 4 Occipital Ctx	30.4
Control 2 Hippo	41.5	AD 5 Occipital Ctx	36.3
Control 4 Hippo	8.4	AD 6 Occipital Ctx	70.2
Control (Path) 3 Hippo	8.1	Control 1 Occipital Ctx	7.5
AD 1 Temporal Ctx	33.7	Control 2 Occipital Ctx	79.6
AD 2 Temporal Ctx	51.8	Control 3 Occipital Ctx	28.5
AD 3 Temporal Ctx	19.3	Control 4 Occipital Ctx	8.8
AD 4 Temporal Ctx	46.0	Control (Path) 1 Occipital Ctx	68.3
AD 5 Inf Temporal Ctx	69.7	Control (Path) 2 Occipital Ctx	16.6
AD 5 SupTemporal Ctx	32.1	Control (Path) 3 Occipital Ctx	4.6
AD 6 Inf Temporal Ctx	48.3	Control (Path) 4 Occipital Ctx	28.9
AD 6 Sup Temporal Ctx	59.0	Control 1 Parietal Ctx	13.7
Control 1 Temporal Ctx	7.7	Control 2 Parietal Ctx	42.0
Control 2 Temporal Ctx	36.6	Control 3 Parietal Ctx	18.6
Control 3 Temporal Ctx	25.7	Control (Path) 1 Parietal Ctx	66.4

Control 4 Temporal Ctx	11.2	Control (Path) 2 Parietal Ctx	31.4
Control (Path) 1 Temporal Ctx	59.9	Control (Path) 3 Parietal Ctx	6.7
Control (Path) 2 Temporal Ctx	51.8	Control (Path) 4 Parietal Ctx	48.3

Table QC. Panel 1.1

Tissue Name	Rel. Exp.(%) Ag399, Run 109660137	Tissue Name	Rel. Exp.(%) Ag399, Run 109660137
Adrenal gland	3.6	Renal ca. UO-31	2.2
Bladder	4.1	Renal ca. RXF 393	0.1
Brain (amygdala)	1.5	Liver	7.5
Brain (cerebellum)	16.6	Liver (fetal)	4.6
Brain (hippocampus)	4.9	Liver ca. (hepatoblast) HepG2	0.0
Brain (substantia nigra)	17.2	Lung	1.7
Brain (thalamus)	6.5	Lung (fetal)	5.9
Cerebral Cortex	15.9	Lung ca. (non-s.cell) HOP-62	100.0
Brain (fetal)	4.8	Lung ca. (large cell)NCI-H460	3.4
Brain (whole)	12.9	Lung ca. (non-s.cell) NCI-H23	1.7
glio/astro U-118-MG	0.8	Lung ca. (non-s.cl) NCI-H522	3.9
astrocytoma SF-539	1.9	Lung ca. (non-sm. cell) A549	3.9
astrocytoma SNB-75	1.1	Lung ca. (s.cell var.) SHP-77	0.3
astrocytoma SW1783	0.2	Lung ca. (small cell) LX-1	4.2
glioma U251	1.1	Lung ca. (small cell) NCI-H69	0.6
glioma SF-295	1.4	Lung ca. (squam.) SW 900	0.8
glioma SNB-19	4.8	Lung ca. (squam.) NCI-H596	1.5
glio/astro U87-MG	3.1	Lymph node	1.6
neuro*; met SK-N-AS	1.3	Spleen	1.1
Mammary gland	4.4	Thymus	1.4
Breast ca. BT-549	0.4	Ovary	3.1
Breast ca. MDA-N	1.6	Ovarian ca. IGROV-1	2.9
Breast ca.* (pl.ef) T47D	8.8	Ovarian ca. OVCAR-3	2.1
Breast ca.* (pl.ef) MCF-7	1.2	Ovarian ca. OVCAR-4	3.1
Breast ca.* (pl.ef) MDA-MB-231	0.4	Ovarian ca. OVCAR-5	3.6
Small intestine	10.8	Ovarian ca. OVCAR-8	3.8
Colorectal	3.6	Ovarian ca.* (ascites) SK-OV-3	0.8

Colon ca. HT29	0.5	Pancreas	21.5
Colon ca. CaCo-2	0.9	Pancreatic ca. CAPAN 2	0.2
Colon ca. HCT-15	1.3	Pituitary gland	19.9
Colon ca. HCT-116	0.6	Placenta	3.2
Colon ca. HCC-2998	3.0	Prostate	7.9
Colon ca. SW480	0.3	Prostate ca.* (bone met) PC-3	8.1
Colon ca.* SW620 (SW480 met)	1.0	Salivary gland	8.1
Stomach	5.9	Trachea	2.1
Gastric ca. (liver met) NCI-N87	4.5	Spinal cord	4.1
Heart	34.6	Testis	4.5
Skeletal muscle (Fetal)	1.8	Thyroid	13.6
Skeletal muscle	36.9	Uterus	2.7
Endothelial cells	11.3	Melanoma M14	1.1
Heart (Fetal)	14.5	Melanoma LOX IMVI	0.0
Kidney	22.2	Melanoma UACC-62	1.6
Kidney (fetal)	8.6	Melanoma SK-MEL-28	13.8
Renal ca. 786-0	0.7	Melanoma* (met) SK- MEL-5	0.9
Renal ca. A498	0.3	Melanoma Hs688(A).T	0.2
Renal ca. ACHN	0.9	Melanoma* (met) Hs688(B).T	0.2
Renal ca. TK-10	1.7		

Table QD. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag399, Run 119216109	Tissue Name	Rel. Exp.(%) Ag399, Run 119216109
Endothelial cells	3.1	Renal ca. 786-0	1.5
Heart (Fetal)	13.9	Renal ca. A498	0.7
Pancreas	54.0	Renal ca. RXF 393	0.6
Pancreatic ca. CAPAN 2	0.6	Renal ca. ACHN	1.2
Adrenal Gland	27.5	Renal ca. UO-31	3.6
Thyroid	64.6	Renal ca. TK-10	2.4
Salivary gland	22.5	Liver	24.3
Pituitary gland	100.0	Liver (fetal)	7.3
Brain (fetal)	25.5	Liver ca. (hepatoblast) HepG2	3.2
Brain (whole)	51.4	Lung	8.4
Brain (amygdala)	12.7	Lung (fetal)	20.0
Brain (cerebellum)	10.5	Lung ca. (small cell) LX- 1	3.8
Brain (hippocampus)	30.1	Lung ca. (small cell) NCI-H69	1.2
Brain (thalamus)	18.6	Lung ca. (s.cell var.) SHP-77	0.6
Cerebral Cortex	30.4	Lung ca. (large cell)NCI- H460	6.0

Spinal cord	11.1	Lung ca. (non-sm. cell) A549	8.0
glio/astro U87-MG	6.8	Lung ca. (non-s.cell) NCI-H23	2.5
glio/astro U-118-MG	1.6	Lung ca. (non-s.cell) HOP-62	40.3
astrocytoma SW1783	0.6	Lung ca. (non-s.cl) NCI-H522	6.7
neuro*; met SK-N-AS	3.7	Lung ca. (squam.) SW 900	1.7
astrocytoma SF-539	1.5	Lung ca. (squam.) NCI-H596	3.4
astrocytoma SNB-75	0.7	Mammary gland	21.9
glioma SNB-19	9.9	Breast ca.* (pl.ef) MCF-7	3.5
glioma U251	3.1	Breast ca.* (pl.ef) MDA-MB-231	1.3
glioma SF-295	4.5	Breast ca.* (pl. ef) T47D	27.0
Heart	41.2	Breast ca. BT-549	1.9
Skeletal Muscle	85.3	Breast ca. MDA-N	3.7
Bone marrow	1.7	Ovary	3.4
Thymus	3.7	Ovarian ca. OVCAR-3	6.6
Spleen	5.4	Ovarian ca. OVCAR-4	3.8
Lymph node	8.7	Ovarian ca. OVCAR-5	9.4
Colorectal Tissue	1.3	Ovarian ca. OVCAR-8	7.3
Stomach	24.1	Ovarian ca. IGROV-1	8.7
Small intestine	30.4	Ovarian ca. (ascites) SK-OV-3	2.4
Colon ca. SW480	0.4	Uterus	10.2
Colon ca.* SW620 (SW480 met)	4.0	Placenta	12.7
Colon ca. HT29	2.1	Prostate	30.8
Colon ca. HCT-116	1.1	Prostate ca.* (bone met) PC-3	15.7
Colon ca. CaCo-2	3.0	Testis	21.6
Colon ca. Tissue (ODO3866)	0.3	Melanoma Hs688(A).T	0.3
Colon ca. HCC-2998	6.2	Melanoma* (met) Hs688(B).T	0.4
Gastric ca.* (liver met) NCI-N87	15.4	Melanoma UACC-62	3.4
Bladder	5.0	Melanoma M14	1.7
Trachea	6.7	Melanoma LOX IMVI	0.1
Kidney	16.7	Melanoma* (met) SK-MEL-5	2.2
Kidney (fetal)	28.7		

Table QE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag399, Run 165678157	Tissue Name	Rel. Exp.(%) Ag399, Run 165678157
Liver adenocarcinoma	5.4	Kidney (fetal)	19.1

Pancreas	10.2	Renal ca. 786-0	3.1
Pancreatic ca. CAPAN 2	6.4	Renal ca. A498	8.0
Adrenal gland	18.3	Renal ca. RXF 393	7.3
Thyroid	20.4	Renal ca. ACHN	1.5
Salivary gland	17.8	Renal ca. UO-31	11.4
Pituitary gland	26.1	Renal ca. TK-10	2.4
Brain (fetal)	24.8	Liver	14.4
Brain (whole)	94.0	Liver (fetal)	32.1
Brain (amygdala)	80.7	Liver ca. (hepatoblast) HepG2	10.3
Brain (cerebellum)	82.4	Lung	12.6
Brain (hippocampus)	100.0	Lung (fetal)	29.9
Brain (substantia nigra)	37.1	Lung ca. (small cell) LX-1	5.6
Brain (thalamus)	85.9	Lung ca. (small cell) NCI-H69	4.1
Cerebral Cortex	77.4	Lung ca. (s.cell var.) SHP-77	6.0
Spinal cord	28.7	Lung ca. (large cell) NCI-H460	13.3
glio/astro U87-MG	9.0	Lung ca. (non-sm. cell) A549	3.7
glio/astro U-118-MG	24.3	Lung ca. (non-s.cell) NCI-H23	2.6
astrocytoma SW1783	4.3	Lung ca. (non-s.cell) HOP-62	38.4
neuro*; met SK-N-AS	2.9	Lung ca. (non-s.cl) NCI-H522	2.4
astrocytoma SF-539	5.0	Lung ca. (squam.) SW 900	2.9
astrocytoma SNB-75	13.6	Lung ca. (squam.) NCI-H596	3.0
glioma SNB-19	28.5	Mammary gland	14.0
glioma U251	35.6	Breast ca. * (pl.ef) MCF-7	5.4
glioma SF-295	5.9	Breast ca. * (pl.ef) MDA-MB-231	11.4
Heart (fetal)	38.4	Breast ca. * (pl.ef) T47D	14.1
Heart	20.2	Breast ca. BT-549	10.9
Skeletal muscle (fetal)	17.1	Breast ca. MDA-N	1.4
Skeletal muscle	50.7	Ovary	8.1
Bone marrow	4.8	Ovarian ca. OVCAR-3	7.4
Thymus	7.6	Ovarian ca. OVCAR-4	6.3
Spleen	16.2	Ovarian ca. OVCAR-5	5.6
Lymph node	39.5	Ovarian ca. OVCAR-8	9.5
Colorectal	14.6	Ovarian ca. IGROV-1	0.5
Stomach	25.0	Ovarian ca. * (ascites) SK-OV-3	3.9
Small intestine	43.2	Uterus	36.6
Colon ca. SW480	1.6	Placenta	7.4
Colon ca. * SW620(SW480 met)	2.5	Prostate	21.2

Colon ca. HT29	0.4	Prostate ca.* (bone met)PC-3	17.2
Colon ca. HCT-116	3.6	Testis	23.5
Colon ca. CaCo-2	6.5	Melanoma Hs688(A).T	2.8
Colon ca. tissue(ODO3866)	3.6	Melanoma* (met) Hs688(B).T	2.3
Colon ca. HCC-2998	3.1	Melanoma UACC-62	5.5
Gastric ca.* (liver met) NCI-N87	20.2	Melanoma M14	16.6
Bladder	3.7	Melanoma LOX IMVI	0.2
Trachea	15.1	Melanoma* (met) SK-MEL-5	2.0
Kidney	14.8	Adipose	10.7

Table QF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag399, Run 165296356	Tissue Name	Rel. Exp.(%) Ag399, Run 165296356
Secondary Th1 act	36.9	HUVEC IL-1beta	7.7
Secondary Th2 act	40.3	HUVEC IFN gamma	56.3
Secondary Tr1 act	44.1	HUVEC TNF alpha + IFN gamma	45.7
Secondary Th1 rest	24.0	HUVEC TNF alpha + IL4	40.9
Secondary Th2 rest	16.5	HUVEC IL-11	22.1
Secondary Tr1 rest	29.3	Lung Microvascular EC none	68.8
Primary Th1 act	29.9	Lung Microvascular EC TNFalpha + IL-1beta	61.1
Primary Th2 act	24.1	Microvascular Dermal EC none	54.0
Primary Tr1 act	46.7	Microvascular Dermal EC TNFalpha + IL-1beta	45.1
Primary Th1 rest	54.7	Bronchial epithelium TNFalpha + IL1beta	95.9
Primary Th2 rest	34.6	Small airway epithelium none	21.5
Primary Tr1 rest	44.1	Small airway epithelium TNFalpha + IL-1beta	43.5
CD45RA CD4 lymphocyte act	26.1	Coronary artery SMC rest	50.3
CD45RO CD4 lymphocyte act	23.8	Coronary artery SMC TNFalpha + IL-1beta	60.3
CD8 lymphocyte act	23.0	Astrocytes rest	61.1
Secondary CD8 lymphocyte rest	14.5	Astrocytes TNFalpha + IL-1beta	60.7
Secondary CD8 lymphocyte act	16.3	KU-812 (Basophil) rest	18.6
CD4 lymphocyte none	24.3	KU-812 (Basophil) PMA/ionomycin	23.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	20.0	CCD1106 (Keratinocytes) none	23.7
LAK cells rest	31.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	30.4
LAK cells IL-2	24.8	Liver cirrhosis	10.2

LAK cells IL-2+IL-12	25.5	Lupus kidney	9.9
LAK cells IL-2+IFN gamma	65.5	NCI-H292 none	40.6
LAK cells IL-2+ IL-18	41.5	NCI-H292 IL-4	42.0
LAK cells PMA/ionomycin	27.2	NCI-H292 IL-9	47.3
NK Cells IL-2 rest	29.1	NCI-H292 IL-13	38.7
Two Way MLR 3 day	56.6	NCI-H292 IFN gamma	28.9
Two Way MLR 5 day	18.3	HPAEC none	53.2
Two Way MLR 7 day	10.5	HPAEC TNF alpha + IL-1 beta	51.1
PBMC rest	12.1	Lung fibroblast none	43.8
PBMC PWM	75.3	Lung fibroblast TNF alpha + IL-1 beta	58.2
PBMC PHA-L	24.1	Lung fibroblast IL-4	67.4
Ramos (B cell) none	41.2	Lung fibroblast IL-9	47.6
Ramos (B cell) ionomycin	88.9	Lung fibroblast IL-13	30.8
B lymphocytes PWM	100.0	Lung fibroblast IFN gamma	35.6
B lymphocytes CD40L and IL-4	87.1	Dermal fibroblast CCD1070 rest	57.4
EOL-1 dbcAMP	17.6	Dermal fibroblast CCD1070 TNF alpha	81.8
EOL-1 dbcAMP PMA/ionomycin	40.1	Dermal fibroblast CCD1070 IL-1 beta	34.6
Dendritic cells none	38.7	Dermal fibroblast IFN gamma	24.3
Dendritic cells LPS	17.7	Dermal fibroblast IL-4	47.3
Dendritic cells anti-CD40	20.6	IBD Colitis 2	4.5
Monocytes rest	27.2	IBD Crohn's	8.8
Monocytes LPS	7.9	Colon	44.8
Macrophages rest	40.1	Lung	60.7
Macrophages LPS	19.2	Thymus	81.2
HUVEC none	22.4	Kidney	77.9
HUVEC starved	51.1		

- CNS\_neurodegeneration\_v1.0 Summary:** Ag399 This panel confirms the expression of the NOV28a gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between
- 5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

- Panel 1.1 Summary:** Ag399 Highest expression of the NOV28a gene is seen in a lung cancer (non-s.cell) cell line HOP-62 (CT=21). Therefore, expression of this gene can be
- 10 used in distinguishing this sample from other samples in the panel. The NOV28a gene encodes a laminin-type EGF-like protein, which belongs to the laminin family. Laminins are the major noncollagenous components of basement membranes that mediate cell adhesion, growth



migration, and differentiation ( Please see Ref. 1 in panel 1.4). Therefore, the moderate to high expression of this gene in samples throughout this panel suggests the possibility of a wider role of this gene product in cell adhesion, growth migration, and differentiation.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at significant levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression (Beck et al., FASEB J. 4: 148-160, 1990).

**Panel 1.2 Summary:** Ag399 Highest expression of the NOV28a gene is seen in the pituitary gland (CT=22). Therefore, expression of this gene can be used in distinguishing this sample from other samples in the panel. In addition, moderate to high expression of this gene is seen samples throughout this panel suggesting the possibility of a wider role of this gene product in cell adhesion, growth migration, and differentiation.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at significant levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 1.3D Summary:** Ag399 Highest expression of the NOV28a gene is detected in brain (hippocampus) sample (CT=29). High expression of this gene is also seen throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, cerebral cortex, spinal cord and glioma cells. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression. In addition, expression of this gene can be used to distinguish the brain

derived tissue samples from other samples used in this panel. The NOV28a gene encodes a laminin-type EGF-like protein, which belongs to the laminin family. Laminins are the major noncollagenous components of basement membranes that mediate cell adhesion, growth migration, and differentiation (Beck et al., 1990). Normal brain cells can produce laminin, fibronectin and collagen type IV when confronted by invading glioma cells. laminin also stimulates cell migration of several human glioma cell lines in vitro (Tysnes et al., Invasion Metastasis 17(5):270-80, 1997).

Low levels of expression of NOV28a gene is also observed in almost all the samples used in this panel suggesting the possibility of a wider role of this gene product in cell adhesion, growth migration, and differentiation.

Among the tissue with metabolic function, this gene is expressed at low to moderate levels in a number of tissues, including adipose, adrenal gland, gastrointestinal tract, pancreas, skeletal muscle and thyroid. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4D Summary:** Ag399 NOV28a codes for laminin-type EGF-like protein, with highest expression in B lymphocytes activated with PWM (CT=30). In addition, this gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General\_screening\_panel\_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

#### **R. NOV29a: polycystic kidney disease 1 protein**

Expression of gene NOV29a was assessed using the primer-probe set Ag3519, described in Table RA. Results of the RTQ-PCR runs are shown in Tables RB, RC and RD.

**Table RA.** Probe Name Ag3519

Primers	Sequences	Length	Start Position
Forward	5'-cacaaatggaactgtgtttgc-3' (SEQ ID NO:257)	21	1134
Probe	TET-5'-cacagacacagacattacattacagctg-3'-TAMRA (SEQ ID NO:258)	29	1155
Reverse	5'-tccaggggtattgtttcctt-3' (SEQ ID NO:259)	20	1189

**Table RB.** CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3519, Run 210630118	Tissue Name	Rel. Exp.(%) Ag3519, Run 210630118
AD 1 Hippo	5.4	Control (Path) 3 Temporal Ctx	10.7
AD 2 Hippo	43.2	Control (Path) 4 Temporal Ctx	62.0
AD 3 Hippo	7.7	AD 1 Occipital Ctx	17.8
AD 4 Hippo	6.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	90.1	AD 3 Occipital Ctx	8.5
AD 6 Hippo	58.6	AD 4 Occipital Ctx	23.7
Control 2 Hippo	11.9	AD 5 Occipital Ctx	17.6
Control 4 Hippo	14.1	AD 6 Occipital Ctx	31.9
Control (Path) 3 Hippo	4.8	Control 1 Occipital Ctx	4.9
AD 1 Temporal Ctx	21.8	Control 2 Occipital Ctx	43.8
AD 2 Temporal Ctx	44.1	Control 3 Occipital Ctx	39.5
AD 3 Temporal Ctx	3.6	Control 4 Occipital Ctx	11.0
AD 4 Temporal Ctx	50.0	Control (Path) 1 Occipital Ctx	66.0
AD 5 Inf Temporal Ctx	57.0	Control (Path) 2 Occipital Ctx	22.5
AD 5 Sup Temporal Ctx	24.1	Control (Path) 3 Occipital Ctx	8.1
AD 6 Inf Temporal Ctx	74.7	Control (Path) 4 Occipital Ctx	33.2
AD 6 Sup Temporal Ctx	100.0	Control 1 Parietal Ctx	21.2
Control 1 Temporal Ctx	15.9	Control 2 Parietal Ctx	42.6
Control 2 Temporal Ctx	20.4	Control 3 Parietal Ctx	17.0
Control 3 Temporal Ctx	11.7	Control (Path) 1 Parietal Ctx	45.4
Control 3 Temporal Ctx	8.2	Control (Path) 2 Parietal Ctx	44.8
Control (Path) 1 Temporal Ctx	52.9	Control (Path) 3 Parietal Ctx	14.6
Control (Path) 2 Temporal Ctx	31.9	Control (Path) 4 Parietal Ctx	80.7

5 **Table RC.** General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3519, Run 216863023	Tissue Name	Rel. Exp.(%) Ag3519, Run 216863023
-------------	------------------------------------	-------------	------------------------------------

Adipose	6.6	Renal ca. TK-10	1.9
Melanoma* Hs688(A).T	0.7	Bladder	10.7
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	11.6
Melanoma* M14	3.9	Gastric ca. KATO III	0.9
Melanoma* LOXIMVI	0.9	Colon ca. SW-948	1.4
Melanoma* SK-MEL-5	15.8	Colon ca. SW480	3.4
Squamous cell carcinoma SCC-4	0.8	Colon ca.* (SW480 met) SW620	2.9
Testis Pool	11.6	Colon ca. HT29	0.6
Prostate ca.* (bone met) PC-3	1.8	Colon ca. HCT-116	3.5
Prostate Pool	5.3	Colon ca. CaCo-2	11.2
Placenta	7.9	Colon cancer tissue	12.8
Uterus Pool	6.3	Colon ca. SW1116	0.9
Ovarian ca. OVCAR-3	3.2	Colon ca. Colo-205	4.8
Ovarian ca. SK-OV-3	7.6	Colon ca. SW-48	2.0
Ovarian ca. OVCAR-4	0.4	Colon Pool	8.2
Ovarian ca. OVCAR-5	31.6	Small Intestine Pool	14.1
Ovarian ca. IGROV-1	0.0	Stomach Pool	22.1
Ovarian ca. OVCAR-8	0.5	Bone Marrow Pool	3.8
Ovary	2.1	Fetal Heart	45.4
Breast ca. MCF-7	0.9	Heart Pool	29.9
Breast ca. MDA-MB- 231	2.1	Lymph Node Pool	6.2
Breast ca. BT 549	3.4	Fetal Skeletal Muscle	17.2
Breast ca. T47D	100.0	Skeletal Muscle Pool	34.2
Breast ca. MDA-N	3.6	Spleen Pool	6.0
Breast Pool	26.1	Thymus Pool	14.4
Trachea	19.2	CNS cancer (glio/astro) U87-MG	0.3
Lung	1.6	CNS cancer (glio/astro) U- 118-MG	1.8
Fetal Lung	5.4	CNS cancer (neuro;met) SK-N-AS	2.2
Lung ca. NCI-N417	0.4	CNS cancer (astro) SF-539	0.9
Lung ca. LX-1	2.0	CNS cancer (astro) SNB-75	3.0
Lung ca. NCI-H146	1.0	CNS cancer (glio) SNB-19	1.0
Lung ca. SHP-77	0.2	CNS cancer (glio) SF-295	5.6
Lung ca. A549	2.3	Brain (Amygdala) Pool	5.8
Lung ca. NCI-H526	0.2	Brain (cerebellum)	8.7
Lung ca. NCI-H23	6.2	Brain (fetal)	7.5
Lung ca. NCI-H460	36.9	Brain (Hippocampus) Pool	5.0
Lung ca. HOP-62	3.4	Cerebral Cortex Pool	6.3
Lung ca. NCI-H522	1.6	Brain (Substantia nigra) Pool	11.0
Liver	1.7	Brain (Thalamus) Pool	10.3
Fetal Liver	2.8	Brain (whole)	21.0
Liver ca. HepG2	1.0	Spinal Cord Pool	8.2

Kidney Pool	21.6	Adrenal Gland	3.5
Fetal Kidney	4.3	Pituitary gland Pool	3.0
Renal ca. 786-0	0.0	Salivary Gland	1.7
Renal ca. A498	0.4	Thyroid (female)	3.8
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.3
Renal ca. UO-31	0.4	Pancreas Pool	4.5

Table RD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3519, Run 166407136	Tissue Name	Rel. Exp.(%) Ag3519, Run 166407136
Secondary Th1 act	0.9	HUVEC IL-1beta	16.5
Secondary Th2 act	0.9	HUVEC IFN gamma	32.1
Secondary Tr1 act	1.7	HUVEC TNF alpha + IFN gamma	5.2
Secondary Th1 rest	0.4	HUVEC TNF alpha + IL4	72.7
Secondary Th2 rest	0.5	HUVEC IL-11	30.6
Secondary Tr1 rest	0.3	Lung Microvascular EC none	2.8
Primary Th1 act	0.5	Lung Microvascular EC TNFalpha + IL-1beta	4.4
Primary Th2 act	3.1	Microvascular Dermal EC none	4.1
Primary Tr1 act	1.6	Microvascular Dermal EC TNFalpha + IL-1beta	4.0
Primary Th1 rest	1.5	Bronchial epithelium TNFalpha + IL1beta	3.3
Primary Th2 rest	0.9	Small airway epithelium none	1.9
Primary Tr1 rest	1.3	Small airway epithelium TNFalpha + IL-1beta	11.2
CD45RA CD4 lymphocyte act	2.3	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	3.7	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	3.7	Astrocytes rest	2.0
Secondary CD8 lymphocyte rest	3.2	Astrocytes TNFalpha + IL-1beta	4.1
Secondary CD8 lymphocyte act	0.8	KU-812 (Basophil) rest	0.4
CD4 lymphocyte none	2.0	KU-812 (Basophil) PMA/ionomycin	1.9
2ry Th1/Th2/Tr1_anti- CD95 CH11	1.6	CCD1106 (Keratinocytes) none	2.3
LAK cells rest	3.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	14.1
LAK cells IL-2	5.8	Liver cirrhosis	14.8
LAK cells IL-2+IL-12	7.6	Lupus kidney	7.5
LAK cells IL-2+IFN gamma	5.9	NCI-H292 none	2.0
LAK cells IL-2+ IL-18	6.2	NCI-H292 IL-4	1.4
LAK cells PMA/ionomycin	2.8	NCI-H292 IL-9	4.1
NK Cells IL-2 rest	2.2	NCI-H292 IL-13	1.4

Two Way MLR 3 day	8.5	NCI-H292 IFN gamma	0.9
Two Way MLR 5 day	3.3	HPAEC none	50.7
Two Way MLR 7 day	1.5	HPAEC TNF alpha + IL-1 beta	100.0
PBMC rest	0.8	Lung fibroblast none	0.9
PBMC PWM	5.3	Lung fibroblast TNF alpha + IL-1 beta	0.4
PBMC PHA-L	3.4	Lung fibroblast IL-4	1.4
Ramos (B cell) none	3.3	Lung fibroblast IL-9	1.7
Ramos (B cell) ionomycin	0.9	Lung fibroblast IL-13	2.9
B lymphocytes PWM	4.3	Lung fibroblast IFN gamma	1.2
B lymphocytes CD40L and IL-4	5.4	Dermal fibroblast CCD1070 rest	1.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	2.2
EOL-1 dbcAMP PMA/ionomycin	2.6	Dermal fibroblast CCD1070 IL-1 beta	0.3
Dendritic cells none	9.7	Dermal fibroblast IFN gamma	0.1
Dendritic cells LPS	5.2	Dermal fibroblast IL-4	1.6
Dendritic cells anti-CD40	17.3	IBD Colitis 2	6.7
Monocytes rest	0.9	IBD Crohn's	2.2
Monocytes LPS	15.5	Colon	15.6
Macrophages rest	27.0	Lung	12.8
Macrophages LPS	18.8	Thymus	6.9
HUVEC none	22.8	Kidney	7.9
HUVEC starved	37.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3519 This panel confirms the expression of the NOV29a gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between

5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3519 Expression of NOV29a is highest in one of the breast cancer T47D cell line (CT=29). Therefore, expression of this gene

10 may be used to distinguish this sample from the other samples on this panel. In addition, low to moderate expression of this gene is detected in large number of samples used in this panel. Therefore, this gene may be playing an important role in cellular function.

In addition, this gene is expressed at moderate levels (CTs=31-33) in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra,

15 thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Moderate expression of this gene is detected in Kidney sample (CT=31). This gene codes for protein similar to polycystic kidney disease (PKD) protein, which is thought to function as part of a multiprotein membrane-spanning complex involved in cell-cell or cell-matrix interactions. Mutations in either of 2 different PKD genes (PKD1 or PKD2) give rise to Autosomal dominant polycystic kidney disease (ADPKD). ADPKD is a major, inherited disorder that is characterized by the growth of large, fluid-filled cysts from the tubules and collecting ducts of affected kidneys, and by a number of extrarenal manifestations including liver and pancreatic cysts, hypertension, heart valve defects, and cerebral and aortic aneurysms (Ref. 1). Therefore, therapeutic modulation of this gene or its protein product may be beneficial in the treatment of ADPKD (Calvet and Grantham, Semin Nephrol 21(2):107-23, 2001).

**Panel 4D Summary:** Ag3519 Low to moderate expression of NOV29a gene is detected in large number of samples used in this panel. Interestingly, expression in LPS stimulated monocytes (CT=32) is higher than in resting monocytes (CT=36). treatment of resting monocytes (CT=36) with LPS stimulated the expression this gene (CT=32). Therefore, expression of this gene may be used to distinguish between these two samples. Highest expression of this gene is seen in TNFalpha + IL-1beta treated HPAEC (CT=29.4). Based on expression in this panel, therapeutic modulation of this gene or its protein product may be beneficial in the treatment of general autoimmunity, rheumatoid disease, asthma, and B-cell disorders.

#### S. NOV30a: POLYCYSTIN 2

Expression of gene NOV30a was assessed using the primer-probe set Ag3522, described in Table SA.

30 Table SA. Probe Name Ag3522

Primers	Sequences	Length	Start Position
Forward	5'-aacttccaagctgttcaaggat-3' (SEQ ID NO:260)	22	1732
Probe	TET-5'-aatgaacaaattatccgccttctctgg-3'-TAMRA (SEQ ID NO:261)	26	1764
Reverse	5'-agcttcactgtggacaggagta-3' (SEQ ID NO:262)	22	1790

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3522 Expression of NOV30a gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**General\_screening\_panel\_v1.4 Summary:** Ag3522 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**Panel 4.1D Summary:** Ag3522 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run.

#### **T. NOV31a: SLIT-like protein**

Expression of gene NOV31a was assessed using the primer-probe sets Ag907 and Ag1925, described in Tables TA and TB. Results of the RTQ-PCR runs are shown in Tables TC, TD, TE and TF.

**Table TA. Probe Name Ag907**

Primers	Sequences	Length	Start Position
Forward	5'-aaagctccagcgtgttgag-3' (SEQ ID NO:263)	19	516
Probe	TET-5'-acctcgatcttgccgaccaggtt-3'-TAMRA (SEQ ID NO:264)	23	468
Reverse	5'-gagattctgcagctgagcaa-3' (SEQ ID NO:265)	20	447

**Table TB. Probe Name Ag1925**

Primers	Sequences	Length	Start Position
Forward	5'-aaagctccagcgtgttgag-3' (SEQ ID NO:266)	19	516
Probe	TET-5'-acctcgatcttgccgaccaggtt-3'-TAMRA (SEQ ID NO:267)	23	468
Reverse	5'-gagattctgcagctgagcaa-3' (SEQ ID NO:268)	20	447

**Table TC. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag907, Run 224758723	Tissue Name	Rel. Exp.(%) Ag907, Run 224758723
AD 1 Hippo	20.3	Control (Path) 3 Temporal Ctx	12.6
AD 2 Hippo	37.6	Control (Path) 4 Temporal Ctx	37.1
AD 3 Hippo	10.2	AD 1 Occipital Ctx	19.8
AD 4 Hippo	13.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	11.3
AD 6 Hippo	39.5	AD 4 Occipital Ctx	16.5
Control 2 Hippo	27.4	AD 5 Occipital Ctx	46.7
Control 4 Hippo	15.4	AD 6 Occipital Ctx	22.4
Control (Path) 3 Hippo	10.7	Control 1 Occipital Ctx	8.4
AD 1 Temporal Ctx	16.3	Control 2 Occipital Ctx	62.0
AD 2 Temporal Ctx	27.2	Control 3 Occipital Ctx	25.3
AD 3 Temporal Ctx	8.8	Control 4 Occipital Ctx	14.1



AD 4 Temporal Ctx	19.3	Control (Path) 1 Occipital Ctx	70.2
AD 5 Inf Temporal Ctx	82.4	Control (Path) 2 Occipital Ctx	15.3
AD 5 Sup Temporal Ctx	44.4	Control (Path) 3 Occipital Ctx	8.4
AD 6 Inf Temporal Ctx	46.3	Control (Path) 4 Occipital Ctx	31.2
AD 6 Sup Temporal Ctx	50.0	Control 1 Parietal Ctx	11.7
Control 1 Temporal Ctx	9.5	Control 2 Parietal Ctx	48.3
Control 2 Temporal Ctx	37.4	Control 3 Parietal Ctx	17.0
Control 3 Temporal Ctx	19.1	Control (Path) 1 Parietal Ctx	66.0
Control 3 Temporal Ctx	17.8	Control (Path) 2 Parietal Ctx	24.3
Control (Path) 1 Temporal Ctx	52.5	Control (Path) 3 Parietal Ctx	9.2
Control (Path) 2 Temporal Ctx	31.6	Control (Path) 4 Parietal Ctx	48.3

Table TD. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag907, Run 119452094	Rel. Exp.(%) Ag907, Run 125218394	Tissue Name	Rel. Exp.(%) Ag907, Run 119452094	Rel. Exp.(%) Ag907, Run 125218394
Endothelial cells	0.0	0.0	Renal ca. 786-0	0.0	0.0
Heart (Fetal)	3.9	8.9	Renal ca. A498	0.7	0.1
Pancreas	4.3	0.0	Renal ca. RXF 393	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. ACHN	0.6	0.1
Adrenal Gland	1.4	0.0	Renal ca. UO-31	0.0	0.0
Thyroid	2.0	0.0	Renal ca. TK-10	0.0	0.0
Salivary gland	1.9	0.4	Liver	1.1	0.1
Pituitary gland	52.1	6.4	Liver (fetal)	0.3	0.0
Brain (fetal)	50.3	15.7	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (whole)	100.0	37.9	Lung	1.6	0.2
Brain (amygdala)	40.9	31.2	Lung (fetal)	3.2	0.5
Brain (cerebellum)	26.6	12.5	Lung ca. (small cell) LX-1	2.4	0.4
Brain (hippocampus)	60.7	35.4	Lung ca. (small cell) NCI-H69	0.0	0.0
Brain (thalamus)	49.0	22.5	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Cerebral Cortex	77.4	100.0	Lung ca. (large cell) NCI-H460	1.5	0.3
Spinal cord	31.6	18.0	Lung ca. (non- sm. cell) A549	2.0	0.2
glio/astro U87-MG	0.0	0.0	Lung ca. (non- s.cell) NCI-H23	2.0	1.7

glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	12.7	5.1
neuro*; met SK-N-AS	2.8	0.2	Lung ca. (squam.) SW 900	0.4	0.0
astrocytoma SF-539	0.0	0.0	Lung ca. (squam.) NCI-H596	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Mammary gland	1.7	0.1
glioma SNB-19	0.1	0.0	Breast ca.* (pl.ef) MCF-7	0.1	0.1
glioma U251	0.1	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
glioma SF-295	1.9	0.1	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	0.1	0.0	Breast ca. BT-549	0.0	0.0
Skeletal Muscle	0.3	0.0	Breast ca. MDA-N	2.9	0.3
Bone marrow	0.0	0.0	Ovary	2.7	2.6
Thymus	0.1	0.0	Ovarian ca. OVCAR-3	0.1	0.0
Spleen	1.6	0.1	Ovarian ca. OVCAR-4	0.1	0.0
Lymph node	2.0	0.6	Ovarian ca. OVCAR-5	5.3	0.5
Colorectal Tissue	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Stomach	2.7	1.1	Ovarian ca. IGROV-1	3.8	1.2
Small intestine	4.2	0.7	Ovarian ca. (ascites) SK-OV-3	0.0	0.0
Colon ca. SW480	0.0	0.0	Uterus	7.3	2.2
Colon ca.* SW620 (SW480 met)	0.0	0.0	Placenta	4.6	1.1
Colon ca. HT29	0.0	0.0	Prostate	2.8	0.6
Colon ca. HCT-116	0.0	0.0	Prostate ca.* (bone met) PC-3	0.9	0.0
Colon ca. CaCo-2	0.1	0.0	Testis	4.4	0.4
Colon ca. Tissue (ODO3866)	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. HCC-2998	10.4	4.3	Melanoma* (met) Hs688(B).T	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.6	0.0	Melanoma UACC-62	0.0	0.0
Bladder	0.1	0.0	Melanoma M14	0.6	0.0
Trachea	0.2	0.0	Melanoma LOX IMVI	0.0	0.0
Kidney	0.1	0.0	Melanoma* (met) SK-MEL-5	1.5	0.4

Kidney (fetal)	0.0	17.3			
----------------	-----	------	--	--	--

Table TE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1925, Run 147205814	Tissue Name	Rel. Exp.(%) Ag1925, Run 147205814
Secondary Th1 act	1.1	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.5
Secondary Tr1 act	1.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	1.4	HUVEC TNF alpha + IL4	0.4
Secondary Th2 rest	0.0	HUVEC IL-11	0.7
Secondary Tr1 rest	0.0	Lung Microvascular EC none	100.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	69.7
Primary Th2 act	0.5	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	1.0
Primary Th1 rest	0.9	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	1.2
Primary Tr1 rest	3.4	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	5.1
CD45RO CD4 lymphocyte act	0.6	Coronary artery SMC TNFalpha + IL-1beta	13.0
CD8 lymphocyte act	0.0	Astrocytes rest	84.7
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	28.3
Secondary CD8 lymphocyte act	0.4	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	1.2	KU-812 (Basophil) PMA/ionomycin	1.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.9	CCD1106 (Keratinocytes) none	1.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.7
LAK cells IL-2	0.8	Liver cirrhosis	3.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.7	NCI-H292 none	0.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.9
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	2.3	NCI-H292 IL-13	0.7
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	2.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	5.5
PBMC rest	0.0	Lung fibroblast none	1.2
PBMC PWM	0.6	Lung fibroblast TNF alpha + IL-	0.0

		l beta	
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.7
Ramos (B cell) none	1.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	1.5	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.8	Lung fibroblast IFN gamma	0.6
B lymphocytes CD40L and IL-4	2.1	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.4
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	3.3	IBD Crohn's	0.0
Monocytes LPS	0.6	Colon	28.5
Macrophages rest	0.8	Lung	41.5
Macrophages LPS	0.5	Thymus	3.2
HUVEC none	0.0	Kidney	2.0
HUVEC starved	0.0		

Table TF. Panel CNS\_1

Tissue Name	Rel. Exp.(%) Ag907, Run 171791128	Tissue Name	Rel. Exp.(%) Ag907, Run 171791128
BA4 Control	33.7	BA17 PSP	42.0
BA4 Control2	51.1	BA17 PSP2	23.3
BA4 Alzheimer's2	15.5	Sub Nigra Control	62.4
BA4 Parkinson's	68.8	Sub Nigra Control2	45.7
BA4 Parkinson's2	96.6	Sub Nigra Alzheimer's2	20.6
BA4 Huntington's	35.1	Sub Nigra Parkinson's2	79.6
BA4 Huntington's2	35.1	Sub Nigra Huntington's	67.8
BA4 PSP	20.0	Sub Nigra Huntington's2	36.6
BA4 PSP2	51.8	Sub Nigra PSP2	9.2
BA4 Depression	31.0	Sub Nigra Depression	15.2
BA4 Depression2	21.2	Sub Nigra Depression2	13.4
BA7 Control	54.3	Glob Palladus Control	27.9
BA7 Control2	65.1	Glob Palladus Control2	16.6
BA7 Alzheimer's2	21.8	Glob Palladus Alzheimer's	21.8
BA7 Parkinson's	33.2	Glob Palladus Alzheimer's2	11.4
BA7 Parkinson's2	62.0	Glob Palladus Parkinson's	100.0
BA7 Huntington's	54.7	Glob Palladus Parkinson's2	23.5
BA7 Huntington's2	64.2	Glob Palladus PSP	7.8
BA7 PSP	48.3	Glob Palladus PSP2	21.3

BA7 PSP2	30.6	Glob Palladus Depression	14.6
BA7 Depression	17.7	Temp Pole Control	16.7
BA9 Control	31.0	Temp Pole Control2	57.8
BA9 Control2	66.9	Temp Pole Alzheimer's	15.9
BA9 Alzheimer's	11.0	Temp Pole Alzheimer's2	9.2
BA9 Alzheimer's2	35.8	Temp Pole Parkinson's	51.8
BA9 Parkinson's	46.7	Temp Pole Parkinson's2	53.6
BA9 Parkinson's2	52.9	Temp Pole Huntington's	56.3
BA9 Huntington's	58.6	Temp Pole PSP	6.9
BA9 Huntington's2	43.5	Temp Pole PSP2	7.0
BA9 PSP	32.3	Temp Pole Depression2	25.7
BA9 PSP2	11.7	Cing Gyr Control	69.3
BA9 Depression	12.6	Cing Gyr Control2	48.0
BA9 Depression2	20.4	Cing Gyr Alzheimer's	24.3
BA17 Control	87.1	Cing Gyr Alzheimer's2	21.5
BA17 Control2	69.7	Cing Gyr Parkinson's	47.0
BA17 Alzheimer's2	27.9	Cing Gyr Parkinson's2	42.3
BA17 Parkinson's	87.7	Cing Gyr Huntington's	82.9
BA17 Parkinson's2	99.3	Cing Gyr Huntington's2	42.0
BA17 Huntington's	64.6	Cing Gyr PSP	32.1
BA17 Huntington's2	46.0	Cing Gyr PSP2	7.6
BA17 Depression	39.0	Cing Gyr Depression	11.3
BA17 Depression2	81.2	Cing Gyr Depression2	25.2

- CNS\_neurodegeneration\_v1.0 Summary:** Ag907 This panel confirms the expression of the NOV31a gene at significant levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between
- 5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.2 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

- Panel 1.2 Summary:** Ag907 Two independent experiments with same probe and primer sets produce results that are in excellent agreement, with high expression of the
- 10 NOV31a gene, a Slit homolog, throughout the CNS, including in amygdala, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. The Slits are a family of secreted guidance proteins that can repel neuronal migration and axon growth via interaction with their cellular roundabout receptors, making this an excellent candidate neuronal guidance protein for axons, dendrites and/or growth cones in general (Ref. 1-2). Therapeutic modulation of the
- 15 levels of this protein, or possible signaling via this protein may be of utility in

enhancing/directing compensatory synaptogenesis and fiber growth in the CNS in response to neuronal death (stroke, head trauma), axon lesion (spinal cord injury), or neurodegeneration (Alzheimer's, Parkinson's, Huntington's, vascular dementia or any neurodegenerative disease). Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's  
 5 disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

In addition, low to moderate expression of this gene is also detected in a melanoma, testis, prostate, prostate cancer, placenta, uterus, ovarian cancer, a breast cancer, mammary gland, lung cancer, adult and fetal lung, adult and fetal liver, lymph node, spleen, skeletal muscle, stomach, small intestine, a colon cancer and a renal cancer sample suggesting the  
 10 possibility of a wider role in intercellular signaling.

Among tissues with metabolic or endocrine function, this gene is expressed at low to moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as  
 15 obesity and diabetes (Battye et al., J. Neurosci. 21: 4290-4298, 2001; Itoh et al., Brain Res. Mol. Brain Res. 62: 175-186, 1998).

**Panel 4D Summary:** Ag907 Moderate to high expression of the NOV31a gene is seen in samples derived from colon, lung, astrocytes, coronary artery SMC, and lung microvascular EC cells. Highest expression of this gene is seen in untreated lung  
 20 microvascular EC cells (CT=29.3). Thus, the expression of this gene could be used to distinguish these samples from the other samples in the panel. Furthermore, expression of this gene is decreased in colon samples from patients with IBD colitis and Crohn's disease (CT=40) relative to normal colon (CT=31.1). Therefore, therapeutic modulation of the activity of the SLIT-like protein encoded by this gene may be useful in the treatment of inflammatory  
 25 bowel disease.

Expression of this gene is in TNFalpha + IL-1beta treated astrocytes and to resting astrocytes (CT=29.53; 84.7%). suggests that therapeutic modulation of the activity of the SLIT-like protein encoded by this gene may also be useful in the treatment of CNS inflammatory disease.

30 **Panel CNS\_1 Summary:** Ag907 This panel confirms expression of the NOV31a gene in the brain. Please see Panel 1.2 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

#### NOV32a: TYROSYLPROTEIN SULFOTRANSFERASE-2

Expression of gene NOV32a was assessed using the primer-probe set Ag3408, described in Table UA. Results of the RTQ-PCR runs are shown in Tables UB and UC.

**Table UA.** Probe Name Ag3408

Primers	Sequences	Length	Start Position
Forward	5'-atcctggaggtgactcttaagc-3' (SEQ ID NO:269)	22	431
Probe	TET-5'-ccatgtgctctccaagaaggaccact-3'-TAMRA (SEQ ID NO:270)	26	466
Reverse	5'-gattcaagggaacttgagtctga-3' (SEQ ID NO:271)	22	492

5

**Table UB.** General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3408, Run 216838909	Tissue Name	Rel. Exp.(%) Ag3408, Run 216838909
Adipose	0.3	Renal ca. TK-10	4.0
Melanoma* Hs688(A).T	0.5	Bladder	3.7
Melanoma* Hs688(B).T	0.5	Gastric ca. (liver met.) NCI-N87	7.4
Melanoma* M14	0.4	Gastric ca. KATO III	4.0
Melanoma* LOXIMV1	0.8	Colon ca. SW-948	0.5
Melanoma* SK-MEL-5	3.3	Colon ca. SW480	4.7
Squamous cell carcinoma SCC-4	1.3	Colon ca.* (SW480 met) SW620	2.7
Testis Pool	0.4	Colon ca. HT29	4.5
Prostate ca.* (bone met) PC-3	1.1	Colon ca. HCT-116	6.0
Prostate Pool	0.9	Colon ca. CaCo-2	5.6
Placenta	0.5	Colon cancer tissue	1.9
Uterus Pool	1.9	Colon ca. SW1116	1.3
Ovarian ca. OVCAR-3	2.7	Colon ca. Colo-205	0.2
Ovarian ca. SK-OV-3	2.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	49.7
Ovarian ca. OVCAR-5	22.4	Small Intestine Pool	4.4
Ovarian ca. IGROV-1	1.4	Stomach Pool	1.6
Ovarian ca. OVCAR-8	1.7	Bone Marrow Pool	1.1
Ovary	1.1	Fetal Heart	9.9
Breast ca. MCF-7	1.4	Heart Pool	6.1
Breast ca. MDA-MB-231	2.5	Lymph Node Pool	5.7
Breast ca. BT 549	3.1	Fetal Skeletal Muscle	2.0
Breast ca. T47D	100.0	Skeletal Muscle Pool	4.8
Breast ca. MDA-N	2.4	Spleen Pool	0.6
Breast Pool	4.5	Thymus Pool	2.8
Trachea	0.8	CNS cancer (glio/astro) U87-MG	4.8
Lung	0.3	CNS cancer (glio/astro) U-118-MG	5.6
Fetal Lung	1.7	CNS cancer (neuro;met) SK-N-AS	5.1

Lung ca. NCI-N417	1.0	CNS cancer (astro) SF-539	2.2
Lung ca. LX-1	3.0	CNS cancer (astro) SNB-75	5.9
Lung ca. NCI-H146	2.5	CNS cancer (glio) SNB-19	1.1
Lung ca. SHP-77	2.0	CNS cancer (glio) SF-295	3.1
Lung ca. A549	1.5	Brain (Amygdala) Pool	0.4
Lung ca. NCI-H526	1.1	Brain (cerebellum)	0.7
Lung ca. NCI-H23	4.9	Brain (fetal)	1.6
Lung ca. NCI-H460	2.7	Brain (Hippocampus) Pool	0.3
Lung ca. HOP-62	0.5	Cerebral Cortex Pool	0.5
Lung ca. NCI-H522	6.1	Brain (Substantia nigra) Pool	0.5
Liver	0.0	Brain (Thalamus) Pool	0.5
Fetal Liver	0.5	Brain (whole)	0.2
Liver ca. HepG2	2.4	Spinal Cord Pool	0.9
Kidney Pool	9.6	Adrenal Gland	0.4
Fetal Kidney	5.0	Pituitary gland Pool	0.4
Renal ca. 786-0	1.4	Salivary Gland	0.0
Renal ca. A498	0.9	Thyroid (female)	0.0
Renal ca. ACHN	1.1	Pancreatic ca. CAPAN2	3.7
Renal ca. UO-31	0.3	Pancreas Pool	4.5

Table UC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3408, Run 165296440	Tissue Name	Rel. Exp.(%) Ag3408, Run 165296440
Secondary Th1 act	12.9	HUVEC IL-1beta	7.4
Secondary Th2 act	19.3	HUVEC IFN gamma	10.9
Secondary Tr1 act	7.6	HUVEC TNF alpha + IFN gamma	23.3
Secondary Th1 rest	6.9	HUVEC TNF alpha + IL4	34.9
Secondary Th2 rest	0.0	HUVEC IL-11	5.2
Secondary Tr1 rest	3.1	Lung Microvascular EC none	20.4
Primary Th1 act	32.8	Lung Microvascular EC TNFalpha + IL-1beta	24.1
Primary Th2 act	32.1	Microvascular Dermal EC none	6.9
Primary Tr1 act	42.0	Microvascular Dermal EC TNFalpha + IL-1beta	14.1
Primary Th1 rest	55.9	Bronchial epithelium TNFalpha + IL1beta	18.7
Primary Th2 rest	15.6	Small airway epithelium none	0.0
Primary Tr1 rest	30.4	Small airway epithelium TNFalpha + IL-1beta	40.6
CD45RA CD4 lymphocyte act	11.0	Coronary artery SMC rest	10.9
CD45RO CD4 lymphocyte act	18.8	Coronary artery SMC TNFalpha + IL-1beta	6.3
CD8 lymphocyte act	20.2	Astrocytes rest	3.3
Secondary CD8 lymphocyte rest	1.9	Astrocytes TNFalpha + IL-1beta	3.6
Secondary CD8	15.7	KU-812 (Basophil) rest	7.2



lymphocyte act			
CD4 lymphocyte none	2.3	KU-812 (Basophil) PMA/ionomycin	43.5
2ry Th1/Th2/Tr1_anti- CD95 CH11	6.7	CCD1106 (Keratinocytes) none	24.7
LAK cells rest	8.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.3
LAK cells IL-2	29.5	Liver cirrhosis	23.5
LAK cells IL-2+IL-12	15.3	Lupus kidney	2.8
LAK cells IL-2+IFN gamma	35.4	NCI-H292 none	38.4
LAK cells IL-2+ IL-18	47.3	NCI-H292 IL-4	100.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	60.3
NK Cells IL-2 rest	42.9	NCI-H292 IL-13	25.7
Two Way MLR 3 day	9.7	NCI-H292 IFN gamma	8.1
Two Way MLR 5 day	7.3	HPAEC none	10.9
Two Way MLR 7 day	17.3	HPAEC TNF alpha + IL-1 beta	20.0
PBMC rest	0.0	Lung fibroblast none	6.5
PBMC PWM	42.6	Lung fibroblast TNF alpha + IL- 1 beta	14.0
PBMC PHA-L	21.9	Lung fibroblast IL-4	47.3
Ramos (B cell) none	46.7	Lung fibroblast IL-9	18.4
Ramos (B cell) ionomycin	94.0	Lung fibroblast IL-13	15.0
B lymphocytes PWM	48.3	Lung fibroblast IFN gamma	25.2
B lymphocytes CD40L and IL-4	37.4	Dermal fibroblast CCD1070 rest	82.4
EOL-1 dbcAMP	13.7	Dermal fibroblast CCD1070 TNF alpha	87.1
EOL-1 dbcAMP PMA/ionomycin	8.1	Dermal fibroblast CCD1070 IL- 1 beta	32.3
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	13.1
Dendritic cells LPS	14.7	Dermal fibroblast IL-4	57.0
Dendritic cells anti-CD40	7.2	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	9.5
Macrophages rest	23.3	Lung	16.2
Macrophages LPS	0.0	Thymus	12.3
HUVEC none	11.7	Kidney	14.9
HUVEC starved	33.7		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3408 Expression of this gene is low/undetectable (CTs > 34) across all of the samples on this panel.

- General\_screening\_panel\_v1.4 Summary:** Ag3408 Highest expression of NOV32a is detected in a breast cancer cell line (CT=2). Therefore, expression of this gene may be used to distinguish this sample from other samples on this panel. In addition, moderate expression of this gene is also observed in an ovarian cancer cell line. Hence, therapeutic modulation of

the activity of this gene product may be beneficial in the treatment of breast and ovarian cancers.

This gene is expressed at low to moderate levels in a number of tissues with metabolic or endocrine function, including gastrointestinal tract, pancreas, and skeletal muscle.

- 5 Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

**Panel 4D Summary:** Ag3408 Highest expression of the NOV32a gene is seen in IL-4 treated NCI-H292 cells (CT=31). However, this gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease.

- 10 These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues.

- 15 Interestingly, expression of this gene is stimulated in PWM/PHA-L treated PBMC cells, IL-2/IL-2+IFN gamma/IL-2+ IL-18 treated LAK cells and ionomycin treated Ramos (B-cell) cells. Therefore, small molecules that antagonize the function of this gene product may be useful as therapeutic drugs to reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which T and B cells play a part in the initiation or  
20 progression of the disease process, such as systemic lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

## V. NOV33a: SERINE PROTEASE INHIBITOR

- 25 Expression of gene NOV33a was assessed using the primer-probe set Ag3436, described in Table VA. Results of the RTQ-PCR runs are shown in Table VB.

Table VA. Probe Name Ag3436

Primers	Sequences	Length	Start Position
Forward	5'-cctcagagctgagtggatga-3' (SEQ ID NO:272)	20	593
Probe	TET-5'-ccctttgactcacgtgccaccag-3'-TAMRA (SEQ ID NO:273)	23	615
Reverse	5'-cgctgtgctcatctacaaaga-3' (SEQ ID NO:274)	21	649

Table VB. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3436, Run 166397093	Tissue Name	Rel. Exp.(%) Ag3436, Run 166397093
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0

Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	4.6	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	1.9	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	1.2
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	11.3	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	9.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0

EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL- 1 beta	0.0
Dendritic cells none	2.4	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	2.0	IBD Colitis 2	4.6
Monocytes rest	0.0	IBD Crohn's	2.7
Monocytes LPS	0.0	Colon	42.3
Macrophages rest	0.0	Lung	10.9
Macrophages LPS	10.9	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3436 Expression of NOV33a gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**General\_screening\_panel\_v1.4 Summary:** Ag3436 Expression of NOV33a gene is low/undetectable (CTs > 35) across all of the samples on this panel.

**Panel 4D Summary:** Ag3436 Highest expression of the NOV33a gene is detected in a liver cirrhosis sample (CT=31.8). Thus, expression of this gene can be used to distinguish this sample from other samples in this panel. Furthermore, therapeutic modulation of the expression or function of this gene could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, expression of this gene could also be used for the diagnosis of liver cirrhosis.

Furthermore, low but significant expression of this gene is detected in the colon (CT=33.1). Expression of this gene is decreased in colon samples from patients with IBD colitis and Crohn's disease (CTs>35). Therefore, therapeutic modulation of the activity of the protein encoded by this gene may be useful in the treatment of inflammatory bowel disease. A related serine protease inhibitor, camostat mesilate, has been used to induce and maintain remission in two patients with ulcerative colitis, to whom salicylazosulfapyridine could not be administered due to previous side effects (Senda et al., Intern Med 32(4):350-4, 1993).

#### 20 **W. NOV34a and NOV34b: Fibronectin type III -like**

Expression of gene NOV34a and NOV34b was assessed using the primer-probe set Ag3538, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB and WC. Please note that NOV34b represents a full-length physical clone of the NOV34a gene, validating the prediction of the gene sequence.

25 **Table WA. Probe Name Ag3538**

Primers	Sequences	Length	Start Position
Forward	5'-gttccagcgcatgaagaag-3' (SEQ ID NO:275)	19	390
Probe	TET-5'-acagctcagaccaagatccagctcct-3'-TAMRA (SEQ ID NO:276)	26	415
Reverse	5'-ggtcgagctgttccaacag-3' (SEQ ID NO:277)	19	454

Table WB. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3538, Run 217044748	Tissue Name	Rel. Exp.(%) Ag3538, Run 217044748
Adipose	0.2	Renal ca. TK-10	0.1
Melanoma* Hs688(A).T	0.7	Bladder	0.6
Melanoma* Hs688(B).T	0.1	Gastric ca. (liver met.) NCI-N87	15.0
Melanoma* M14	1.1	Gastric ca. KATO III	6.0
Melanoma* LOXIMV1	0.8	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	1.4	Colon ca. SW480	29.9
Squamous cell carcinoma SCC-4	7.1	Colon ca.* (SW480 met) SW620	0.7
Testis Pool	100.0	Colon ca. HT29	9.3
Prostate ca.* (bone met) PC-3	6.0	Colon ca. HCT-116	5.6
Prostate Pool	0.5	Colon ca. CaCo-2	2.2
Placenta	0.2	Colon cancer tissue	2.1
Uterus Pool	1.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	12.9	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	2.0
Ovarian ca. OVCAR-4	1.5	Colon Pool	1.6
Ovarian ca. OVCAR-5	16.0	Small Intestine Pool	4.2
Ovarian ca. IGROV-1	13.8	Stomach Pool	0.6
Ovarian ca. OVCAR-8	4.8	Bone Marrow Pool	0.6
Ovary	0.5	Fetal Heart	0.7
Breast ca. MCF-7	2.5	Heart Pool	0.3
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	1.8
Breast ca. BT 549	20.9	Fetal Skeletal Muscle	0.0
Breast ca. T47D	21.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	1.0	Spleen Pool	3.8
Breast Pool	2.4	Thymus Pool	3.6
Trachea	8.7	CNS cancer (glio/astro) U87-MG	0.6
Lung	0.7	CNS cancer (glio/astro) U-118-MG	2.2
Fetal Lung	9.1	CNS cancer (neuro;met) SK-N-AS	4.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	3.8	CNS cancer (astro) SNB-75	1.1
Lung ca. NCI-H146	3.2	CNS cancer (glio) SNB-19	23.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	1.2

Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	19.5	Brain (fetal)	9.8
Lung ca. NCI-H460	1.2	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	3.4	Cerebral Cortex Pool	0.9
Lung ca. NCI-H522	0.4	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.9
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.5
Kidney Pool	3.3	Adrenal Gland	0.0
Fetal Kidney	2.2	Pituitary gland Pool	1.4
Renal ca. 786-0	0.0	Salivary Gland	1.9
Renal ca. A498	2.0	Thyroid (female)	3.0
Renal ca. ACHN	10.6	Pancreatic ca. CAPAN2	15.9
Renal ca. UO-31	1.3	Pancreas Pool	2.0

Table WC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3538, Run 166446357	Tissue Name	Rel. Exp.(%) Ag3538, Run 166446357
Secondary Th1 act	0.0	HUVEC IL-1beta	0.6
Secondary Th2 act	1.8	HUVEC IFN gamma	1.3
Secondary Tr1 act	0.4	HUVEC TNF alpha + IFN gamma	0.4
Secondary Th1 rest	1.1	HUVEC TNF alpha + IL4	1.6
Secondary Th2 rest	0.4	HUVEC IL-11	1.3
Secondary Tr1 rest	0.4	Lung Microvascular EC none	1.5
Primary Th1 act	1.8	Lung Microvascular EC TNFalpha + IL-1beta	0.9
Primary Th2 act	1.6	Microvascular Dermal EC none	1.8
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	1.5
Primary Th1 rest	0.6	Bronchial epithelium TNFalpha + IL1beta	1.3
Primary Th2 rest	0.4	Small airway epithelium none	0.1
Primary Tr1 rest	1.2	Small airway epithelium TNFalpha + IL-1beta	1.3
CD45RA CD4 lymphocyte act	0.3	Coronary artery SMC rest	0.5
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.5
CD8 lymphocyte act	0.0	Astrocytes rest	0.5
Secondary CD8 lymphocyte rest	2.0	Astrocytes TNFalpha + IL-1beta	1.3
Secondary CD8 lymphocyte act	1.5	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	1.1	KU-812 (Basophil) PMA/ionomycin	1.1
2ry Th1/Th2/Tr1_anti-	0.6	CCD1106 (Keratinocytes) none	4.3

CD95 CH11			
LAK cells rest	0.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.2
LAK cells IL-2	1.9	Liver cirrhosis	13.6
LAK cells IL-2+IL-12	0.4	Lupus kidney	0.4
LAK cells IL-2+IFN gamma	2.3	NCI-H292 none	1.5
LAK cells IL-2+ IL-18	1.8	NCI-H292 IL-4	0.8
LAK cells PMA/ionomycin	0.7	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.1	NCI-H292 IL-13	0.4
Two Way MLR 3 day	1.3	NCI-H292 IFN gamma	1.8
Two Way MLR 5 day	0.6	HPAEC none	1.5
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	1.3
PBMC rest	0.3	Lung fibroblast none	4.7
PBMC PWM	0.6	Lung fibroblast TNF alpha + IL- 1 beta	2.5
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.3
Ramos (B cell) none	0.2	Lung fibroblast IL-9	0.5
Ramos (B cell) ionomycin	0.1	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.5	Lung fibroblast IFN gamma	0.1
B lymphocytes CD40L and IL-4	1.4	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.5	Dermal fibroblast CCD1070 TNF alpha	0.8
EOL-1 dbcAMP PMA/ionomycin	3.0	Dermal fibroblast CCD1070 IL- 1 beta	0.0
Dendritic cells none	0.5	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	2.3	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.6
Monocytes rest	0.0	IBD Crohn's	0.6
Monocytes LPS	2.6	Colon	100.0
Macrophages rest	1.8	Lung	8.4
Macrophages LPS	3.1	Thymus	0.5
HUVEC none	0.8	Kidney	1.3
HUVEC starved	2.4		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3538 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel.

- General\_screening\_panel\_v1.4 Summary:** Ag3538 Highest expression of the NOV34a gene is detected in sample derived from testis (CT=29.8). Thus, expression of this gene can be used to distinguish this sample from other samples in the panel. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of fertility disorders and hypogonadism.

In addition, significant expression of this gene is seen in pancreatic, CNS, colon, gastric, renal, lung, breast, ovarian and squamous cell carcinoma cell lines. Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of these cancers.

Interestingly, this gene is expressed at much higher levels in fetal (CT = 33) when compared to adult brain samples (CT=36-40). This observation suggests that expression of this gene can be used to distinguish fetal from adult brain.

**Panel 4D Summary:** Ag3538 Highest expression of NOV34a is detected in sample derived from colon (CT=29.78). Thus, expression of this gene can be used to distinguish this sample from other samples in the panel. Furthermore, expression of this gene is decreased in colon samples from patients with IBD colitis and Crohn's disease (CTs>37) relative to normal colon. Therefore, therapeutic modulation of the activity of the GPCR encoded by this gene may be useful in the treatment of inflammatory bowel disease.

#### X. NOV35a: ADIPOPHILIN (ADIPOSE DIFFERENTIATION-RELATED PROTEIN)

Expression of gene NOV35a was assessed using the primer-probe set Ag5733, described in Table XA. Results of the RTQ-PCR runs are shown in Table XB.

Table XA. Probe Name Ag5733

Primers	Sequences	Length	Start Position
Forward	5'-agttgatccacaaccgagtgt-3' (SEQ ID NO:278)	21	83
Probe	TET-5'-ccttggtgagctccacgtatgacct-3'-TAMRA (SEQ ID NO:279)	25	127
Reverse	5'-actgagataggctgaggacatg-3' (SEQ ID NO:280)	22	152

Table XB. General screening panel\_v1.5

Tissue Name	Rel. Exp.(%) Ag5733, Run 245455392	Tissue Name	Rel. Exp.(%) Ag5733, Run 245455392
Adipose	15.8	Renal ca. TK-10	76.8
Melanoma* Hs688(A).T	13.1	Bladder	11.5
Melanoma* Hs688(B).T	25.3	Gastric ca. (liver met.) NCI-N87	5.8
Melanoma* M14	23.7	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	38.4	Colon ca. SW-948	2.3
Melanoma* SK-MEL-5	13.6	Colon ca. SW480	13.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	24.0
Testis Pool	3.3	Colon ca. HT29	2.3



Prostate ca. * (bone met) PC-3	4.2	Colon ca. HCT-116	6.6
Prostate Pool	1.3	Colon ca. CaCo-2	25.9
Placenta	19.1	Colon cancer tissue	14.9
Uterus Pool	3.3	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	4.2	Colon ca. Colo-205	5.2
Ovarian ca. SK-OV-3	0.9	Colon ca. SW-48	12.3
Ovarian ca. OVCAR-4	2.2	Colon Pool	2.2
Ovarian ca. OVCAR-5	0.8	Small Intestine Pool	2.4
Ovarian ca. IGROV-1	0.0	Stomach Pool	3.0
Ovarian ca. OVCAR-8	2.4	Bone Marrow Pool	2.5
Ovary	2.6	Fetal Heart	5.0
Breast ca. MCF-7	0.4	Heart Pool	2.6
Breast ca. MDA-MB-231	12.6	Lymph Node Pool	5.6
Breast ca. BT 549	63.7	Fetal Skeletal Muscle	4.0
Breast ca. T47D	1.9	Skeletal Muscle Pool	34.2
Breast ca. MDA-N	12.6	Spleen Pool	1.4
Breast Pool	2.6	Thymus Pool	3.1
Trachea	3.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.8	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	6.1	CNS cancer (neuro;met) SK-N-AS	6.4
Lung ca. NCI-N417	0.1	CNS cancer (astro) SF-539	12.9
Lung ca. LX-1	20.4	CNS cancer (astro) SNB-75	2.2
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	1.1	CNS cancer (glio) SF-295	0.1
Lung ca. A549	5.6	Brain (Amygdala) Pool	0.6
Lung ca. NCI-H526	0.6	Brain (cerebellum)	1.3
Lung ca. NCI-H23	3.6	Brain (fetal)	2.6
Lung ca. NCI-H460	0.8	Brain (Hippocampus) Pool	0.7
Lung ca. HOP-62	10.2	Cerebral Cortex Pool	0.7
Lung ca. NCI-H522	9.7	Brain (Substantia nigra) Pool	0.7
Liver	21.5	Brain (Thalamus) Pool	0.7
Fetal Liver	25.9	Brain (whole)	5.1
Liver ca. HepG2	100.0	Spinal Cord Pool	1.2
Kidney Pool	5.2	Adrenal Gland	9.7
Fetal Kidney	4.4	Pituitary gland Pool	0.3
Renal ca. 786-0	14.1	Salivary Gland	3.3
Renal ca. A498	77.9	Thyroid (female)	0.8
Renal ca. ACHN	15.9	Pancreatic ca. CAPAN2	1.4
Renal ca. UO-31	18.4	Pancreas Pool	33.2

**General\_screening\_panel\_v1.5 Summary:** Ag5733 Highest expression of NOV35a is detected in sample derived from liver cancer cell line (CT=24). Thus, expression of this

gene can be used to distinguish this sample from other samples in this panel. In addition, high expression of this gene is also associated with renal cancer, melanoma, breast cancer, colon cancer, and lung cancer cell lines. Therefore, therapeutic modulation of this gene product may be beneficial in the treatment of these cancers.

- 5 This gene is expressed at moderate to high levels in a number of tissues with metabolic or endocrine function, including adipose, adrenal gland, gastrointestinal tract, pancreas, skeletal muscle and thyroid. The NOV35a gene codes for adipophilin, which belongs to perilipin family. Perilipin is known to play a role in regulation of triacylglycerol hydrolysis and lipid metabolism of adipose tissue (Ref.1). Therefore, therapeutic modulation of the
- 10 activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

- In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central
- 15 nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression (Tansey et al., Proc Natl Acad Sci U S A 98(11):6494-9, 2001).

#### **Y. NOV37a, NOV37b and NOV37c: Latent transforming growth factor beta**

##### **20 binding protein 1**

Expression of gene NOV37a, NOV37b and NOV37c was assessed using the primer-probe set Ag3596, described in Table YA. Results of the RTQ-PCR runs are shown in Tables YB, YC and YD.

Table YA. Probe Name Ag3596

Primers	Sequences	Length	Start Position
Forward	5'-gatgtatacgaccggctgagt-3' (SEQ ID NO:281)	21	4569
Probe	TET-5'-cgaacaaatagaagaactgatgtctacca-3'-TAMRA (SEQ ID NO:282)	30	4594
Reverse	5'-agatgttcccagcacaatct-3' (SEQ ID NO:283)	21	4624

25

Table YB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3596, Run 211010102	Tissue Name	Rel. Exp.(%) Ag3596, Run 211010102
AD 1 Hippo	16.7	Control (Path) 3 Temporal Ctx	6.1
AD 2 Hippo	35.8	Control (Path) 4 Temporal Ctx	20.9
AD 3 Hippo	9.5	AD 1 Occipital Ctx	15.8

AD 4 Hippo	16.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	29.9	AD 3 Occipital Ctx	4.8
AD 6 Hippo	100.0	AD 4 Occipital Ctx	30.1
Control 2 Hippo	17.1	AD 5 Occipital Ctx	25.2
Control 4 Hippo	33.2	AD 6 Occipital Ctx	19.5
Control (Path) 3 Hippo	11.4	Control 1 Occipital Ctx	7.4
AD 1 Temporal Ctx	26.8	Control 2 Occipital Ctx	29.5
AD 2 Temporal Ctx	32.1	Control 3 Occipital Ctx	10.7
AD 3 Temporal Ctx	8.7	Control 4 Occipital Ctx	17.8
AD 4 Temporal Ctx	38.2	Control (Path) 1 Occipital Ctx	44.4
AD 5 Inf Temporal Ctx	32.3	Control (Path) 2 Occipital Ctx	7.7
AD 5 Sup Temporal Ctx	54.0	Control (Path) 3 Occipital Ctx	1.5
AD 6 Inf Temporal Ctx	43.8	Control (Path) 4 Occipital Ctx	10.6
AD 6 Sup Temporal Ctx	60.7	Control 1 Parietal Ctx	14.6
Control 1 Temporal Ctx	8.7	Control 2 Parietal Ctx	33.2
Control 2 Temporal Ctx	16.4	Control 3 Parietal Ctx	9.2
Control 3 Temporal Ctx	9.3	Control (Path) 1 Parietal Ctx	29.1
Control 3 Temporal Ctx	14.0	Control (Path) 2 Parietal Ctx	30.8
Control (Path) 1 Temporal Ctx	38.4	Control (Path) 3 Parietal Ctx	4.0
Control (Path) 2 Temporal Ctx	20.9	Control (Path) 4 Parietal Ctx	21.5

Table YC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3596, Run 218307094	Tissue Name	Rel. Exp.(%) Ag3596, Run 218307094
Adipose	2.5	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	4.8	Bladder	7.1
Melanoma* Hs688(B).T	10.4	Gastric ca. (liver met.) NCI-N87	1.8
Melanoma* M14	0.0	Gastric ca. KATO III	4.0
Melanoma* LOXIMV1	0.3	Colon ca. SW-948	0.6
Melanoma* SK-MEL-5	1.9	Colon ca. SW480	0.3
Squamous cell carcinoma SCC-4	1.2	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	5.8	Colon ca. HT29	0.4
Prostate ca.* (bone met) PC-3	63.7	Colon ca. HCT-116	0.9
Prostate Pool	8.1	Colon ca. CaCo-2	4.7
Placenta	7.5	Colon cancer tissue	10.9
Uterus Pool	9.5	Colon ca. SW1116	0.4
Ovarian ca. OVCAR-3	12.7	Colon ca. Colo-205	0.0

Ovarian ca. SK-OV-3	8.4	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.7	Colon Pool	20.7
Ovarian ca. OVCAR-5	5.3	Small Intestine Pool	7.3
Ovarian ca. IGROV-1	5.1	Stomach Pool	8.5
Ovarian ca. OVCAR-8	9.2	Bone Marrow Pool	9.5
Ovary	5.5	Fetal Heart	16.4
Breast ca. MCF-7	5.0	Heart Pool	8.1
Breast ca. MDA-MB-231	5.1	Lymph Node Pool	24.5
Breast ca. BT 549	21.2	Fetal Skeletal Muscle	3.7
Breast ca. T47D	14.6	Skeletal Muscle Pool	2.5
Breast ca. MDA-N	0.1	Spleen Pool	1.2
Breast Pool	19.8	Thymus Pool	11.3
Trachea	5.4	CNS cancer (glio/astro) U87-MG	46.3
Lung	3.2	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	18.3	CNS cancer (neuro;met) SK-N-AS	1.4
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF-539	2.7
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	14.6
Lung ca. NCI-H146	0.5	CNS cancer (glio) SNB-19	5.8
Lung ca. SHP-77	0.6	CNS cancer (glio) SF-295	100.0
Lung ca. A549	7.7	Brain (Amygdala) Pool	0.7
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	6.5	Brain (fetal)	2.8
Lung ca. NCI-H460	1.2	Brain (Hippocampus) Pool	2.0
Lung ca. HOP-62	2.2	Cerebral Cortex Pool	1.3
Lung ca. NCI-H522	4.4	Brain (Substantia nigra) Pool	0.9
Liver	0.2	Brain (Thalamus) Pool	1.3
Fetal Liver	9.3	Brain (whole)	1.5
Liver ca. HepG2	0.0	Spinal Cord Pool	1.4
Kidney Pool	22.2	Adrenal Gland	2.1
Fetal Kidney	5.9	Pituitary gland Pool	1.1
Renal ca. 786-0	0.0	Salivary Gland	1.5
Renal ca. A498	0.3	Thyroid (female)	0.4
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.8
Renal ca. UO-31	0.0	Pancreas Pool	14.6

Table YD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3596, Run 169910408	Tissue Name	Rel. Exp.(%) Ag3596, Run 169910408
Secondary Th1 act	0.0	HUVEC IL-1beta	15.6
Secondary Th2 act	0.0	HUVEC IFN gamma	14.6
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	10.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	11.3

Secondary Th2 rest	0.0	HUVEC IL-11	7.2
Secondary Tr1 rest	0.0	Lung Microvascular EC none	13.2
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	8.9
Primary Th2 act	0.0	Microvascular Dermal EC none	29.1
Primary Tr1 act	0.1	Microvascular Dermal EC TNFalpha + IL-1beta	18.7
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	2.4
Primary Th2 rest	0.0	Small airway epithelium none	6.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	1.9
CD45RA CD4 lymphocyte act	5.6	Coronary artery SMC rest	19.6
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	16.8
CD8 lymphocyte act	0.0	Astrocytes rest	7.6
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	7.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.1
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	3.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	1.4
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.0
LAK cells IL-2	0.0	Liver cirrhosis	2.0
LAK cells IL-2+IL-12	0.2	NCI-H292 none	1.8
LAK cells IL-2+IFN gamma	0.1	NCI-H292 IL-4	3.0
LAK cells IL-2+ IL-18	0.1	NCI-H292 IL-9	3.4
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	2.1
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	1.3
Two Way MLR 3 day	0.0	HPAEC none	5.1
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	4.0
Two Way MLR 7 day	0.0	Lung fibroblast none	81.2
PBMC rest	0.2	Lung fibroblast TNF alpha + IL-1 beta	68.3
PBMC PWM	0.0	Lung fibroblast IL-4	59.5
PBMC PHA-L	0.0	Lung fibroblast IL-9	100.0
Ramos (B cell) none	16.0	Lung fibroblast IL-13	81.8
Ramos (B cell) ionomycin	10.7	Lung fibroblast IFN gamma	88.9
B lymphocytes PWM	0.2	Dermal fibroblast CCD1070 rest	20.9
B lymphocytes CD40L and IL-4	0.8	Dermal fibroblast CCD1070 TNF alpha	14.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	13.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	20.4

Dendritic cells none	0.0	Dermal fibroblast IL-4	58.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	53.2
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.1
Monocytes rest	0.1	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	3.2
Macrophages rest	0.0	Lung	13.9
Macrophages LPS	0.0	Thymus	1.8
HUVEC none	14.1	Kidney	5.4
HUVEC starved	14.9		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3596 This panel confirms the expression of the NOV37a gene at low levels in the brain in an independent group of individuals. This gene is found to be upregulated in the temporal cortex of Alzheimer's disease patients. Blockade of this gene product may be useful in the treatment of this disease and decrease neuronal death.

**General\_screening\_panel\_v1.4 Summary:** Ag3596 Highest expression of the NOV37a gene is detected in one of the CNS cancer cell line (CT=26). Thus, expression of this gene can be used to distinguish this sample from other samples in this panel. In addition, significant expression of this gene is also associated with prostate cancer (CT=27). Therefore, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be beneficial in the treatment of CNS cancer or prostate cancer.

In prostatic carcinoma there is immunohistochemical evidence that TGF-beta 1 is produced without the associated-LTBP1 in malignant cells, although TGF beta1-LTBP1 complexes are present in cystectomized prostatic and benign prostatic hyperplastic tissues (Eklov et al., Cancer Res. 53, 3193-3197, 1993).

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at significant levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

**Panel 4.1D Summary:** Ag3596 Highest expression of the NOV37a gene is detected in IL-9 treated lung fibroblast (CT=27.3). In addition, high expression of this gene is detected in TNF alpha + IL-1 beta/IL-4/IL-13/IFN gamma treated as well as untreated lung fibroblast and also in IFN gamma treated and untreated dermal fibroblasts. Thus, expression of this gene can be used to distinguish the lung and dermal fibroblast samples from other samples in this panel. Also, modulation of this gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, psoriasis and idiopathic pulmonary fibrosis (IPF).

Recently, Saika et al. (Saika et al., Graefes Arch Clin Exp Ophthalmol 239(3):234-41, 2001) have shown that LTBP-1, beta 1-LAP and fibrillin-1 co-localize to the ECM of the filtering bleb and of cultured conjunctival fibroblasts. Both conjunctival epithelium and fibroblasts are considered to be the source of TGF beta in healing bleb. ECM secreted by in vivo and in vitro subconjunctival fibroblasts may work as a scavenger or repository of TGF beta.

#### **Z. NOV39a: Urokinase plasminogen activator surface receptor precursor**

Expression of gene NOV39a was assessed using the primer-probe set Ag3134, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB, ZC, ZD, ZE and ZF.

**Table ZA. Probe Name Ag3134**

Primers	Sequences	Length	Start Position
Forward	5'-agctttgagcacacctactttg-3' (SEQ ID NO:284)	22	82
Probe	TET-5'-cccagcatctcctgtcctcatgagt-3'-TAMRA (SEQ ID NO:285)	25	133
Reverse	5'-agagacaggatagcctcaaagc-3' (SEQ ID NO:286)	22	158

**Table ZB. CNS\_neurodegeneration\_v1.0**

Tissue Name	Rel. Exp.(%) Ag3134, Run 209055794	Tissue Name	Rel. Exp.(%) Ag3134, Run 209055794
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	23.3	Control (Path) 4 Temporal Ctx	15.4
AD 3 Hippo	0.5	AD 1 Occipital Ctx	0.0
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	0.0

AD 6 Hippo	38.7	AD 4 Occipital Ctx	15.7
Control 2 Hippo	19.3	AD 5 Occipital Ctx	37.1
Control 4 Hippo	0.8	AD 6 Occipital Ctx	8.5
Control (Path) 3 Hippo	0.2	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	0.0	Control 2 Occipital Ctx	70.2
AD 2 Temporal Ctx	29.9	Control 3 Occipital Ctx	0.3
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	0.0	Control (Path) 1 Occipital Ctx	65.5
AD 5 Inf Temporal Ctx	74.7	Control (Path) 2 Occipital Ctx	3.5
AD 5 Sup Temporal Ctx	29.5	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	36.3	Control (Path) 4 Occipital Ctx	4.5
AD 6 Sup Temporal Ctx	44.8	Control 1 Parietal Ctx	0.9
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	22.7
Control 2 Temporal Ctx	33.0	Control 3 Parietal Ctx	7.3
Control 3 Temporal Ctx	7.0	Control (Path) 1 Parietal Ctx	76.3
Control 3 Temporal Ctx	1.2	Control (Path) 2 Parietal Ctx	9.8
Control (Path) 1 Temporal Ctx	52.1	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	14.3	Control (Path) 4 Parietal Ctx	45.1

Table ZC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3134, Run 165552410	Tissue Name	Rel. Exp.(%) Ag3134, Run 165552410
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	1.3	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.1	Renal ca. UO-31	0.0
Pituitary gland	0.5	Renal ca. TK-10	0.0
Brain (fetal)	13.8	Liver	0.0
Brain (whole)	75.3	Liver (fetal)	0.0
Brain (amygdala)	32.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	29.5	Lung	15.8
Brain (hippocampus)	33.9	Lung (fetal)	4.2
Brain (substantia nigra)	31.2	Lung ca. (small cell) LX-1	0.2
Brain (thalamus)	97.3	Lung ca. (small cell) NCI-H69	0.9
Cerebral Cortex	29.7	Lung ca. (s.cell var.) SHP-77	0.0



Spinal cord	19.5	Lung ca. (large cell)NCI-H460	2.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	2.2
neuro*; met SK-N-AS	0.7	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.1
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.1
glioma SNB-19	5.6	Mammary gland	1.1
glioma U251	4.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.3	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.4
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	2.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	21.3	Ovarian ca. OVCAR-5	0.0
Lymph node	2.0	Ovarian ca. OVCAR-8	11.6
Colorectal	0.1	Ovarian ca. IGROV-1	0.0
Stomach	2.2	Ovarian ca.* (ascites) SK-OV-3	0.4
Small intestine	3.0	Uterus	5.1
Colon ca. SW480	0.8	Placenta	100.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.6
Colon ca. HCT-116	0.0	Testis	0.3
Colon ca. CaCo-2	0.7	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	5.3	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	12.9
Bladder	0.2	Melanoma LOX IMVI	0.0
Trachea	4.3	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table ZD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3134,	Tissue Name	Rel. Exp.(%) Ag3134,
-------------	----------------------	-------------	----------------------

	Run 165910588		Run 165910588
Normal Colon	10.6	Kidney Margin 8120608	1.3
CC Well to Mod Diff (ODO3866)	7.4	Kidney Cancer 8120613	0.7
CC Margin (ODO3866)	0.9	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	0.7	Kidney Cancer 9010320	4.4
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	3.5
CC Mod Diff (ODO3920)	3.1	Normal Uterus	2.4
CC Margin (ODO3920)	2.7	Uterus Cancer 064011	7.5
CC Gr.2 ascend colon (ODO3921)	12.6	Normal Thyroid	0.9
CC Margin (ODO3921)	1.5	Thyroid Cancer 064010	3.3
CC from Partial Hepatectomy (ODO4309) Mets	7.9	Thyroid Cancer A302152	6.1
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	3.7
Colon mets to lung (OD04451-01)	3.8	Normal Breast	15.1
Lung Margin (OD04451-02)	15.1	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	4.9	Breast Cancer (OD04590-01)	4.9
Prostate Cancer (OD04410)	7.1	Breast Cancer Mets (OD04590-03)	6.1
Prostate Margin (OD04410)	6.0	Breast Cancer Metastasis (OD04655-05)	0.2
Prostate Cancer (OD04720-01)	23.5	Breast Cancer 064006	5.3
Prostate Margin (OD04720-02)	15.0	Breast Cancer 1024	25.7
Normal Lung 061010	25.7	Breast Cancer 9100266	6.3
Lung Met to Muscle (ODO4286)	1.0	Breast Margin 9100265	6.1
Muscle Margin (ODO4286)	0.1	Breast Cancer A209073	17.1
Lung Malignant Cancer (OD03126)	9.1	Breast Margin A209073	14.0
Lung Margin (OD03126)	24.8	Normal Liver	0.0
Lung Cancer (OD04404)	85.9	Liver Cancer 064003	0.4
Lung Margin (OD04404)	4.5	Liver Cancer 1025	0.2
Lung Cancer (OD04565)	42.0	Liver Cancer 1026	2.5
Lung Margin (OD04565)	5.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	26.1	Liver Tissue 6004-N	0.1
Lung Margin (OD04237-02)	41.5	Liver Cancer 6005-T	3.1
Ocular Mel Met to Liver (ODO4310)	12.2	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	9.1
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	1.3
Lung Margin (OD04321)	27.2	Bladder Cancer A302173	23.7
Normal Kidney	6.0	Bladder Cancer (OD04718-01)	100.0
Kidney Ca, Nuclear grade 2 (OD04338)	6.5	Bladder Normal Adjacent (OD04718-03)	1.4
Kidney Margin (OD04338)	9.5	Normal Ovary	3.1

Kidney Ca Nuclear grade 1/2 (OD04339)	0.5	Ovarian Cancer 064008	9.6
Kidney Margin (OD04339)	1.9	Ovarian Cancer (OD04768-07)	0.5
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	2.3
Kidney Margin (OD04340)	3.6	Normal Stomach	9.8
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.6
Kidney Margin (OD04348)	8.7	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.7	Gastric Cancer 9060395	5.4
Kidney Margin (OD04622-03)	0.6	Stomach Margin 9060394	0.1
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	10.7
Kidney Margin (OD04450-03)	3.2	Stomach Margin 9060396	0.3
Kidney Cancer 8120607	0.2	Gastric Cancer 064005	2.4

Table ZE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3134, Run 166618735	Tissue Name	Rel. Exp.(%) Ag3134, Run 166618735
Daoy- Medulloblastoma	4.4	Ca Ski- Cervical epidermoid carcinoma (metastasis)	9.4
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	6.9	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megakaryoblast)	0.0
SNB-78- Glioma	0.2	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.6	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.1	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	26.2	JM1- pre-B-cell lymphoma	0.0
Cerebellum	19.3	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	1.1	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	100.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	28.1	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	2.1	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0

NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.7
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	2.6	Hs766T- Pancreatic carcinoma (LN metastasis)	4.1
NCI-H1299- Large cell lung cancer	0.4	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.4
NCI-H727- Lung carcinoid	0.3	SU86.86- Pancreatic carcinoma (liver metastasis)	0.0
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	12.7
LX-1- Small cell lung cancer	6.5	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	3.7	CFPAC-1- Pancreatic ductal adenocarcinoma	6.7
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	1.7
NCI-H716- Colon cancer	0.3	T24- Bladder carcinoma (transitional cell)	0.4
SW-48- Colon adenocarcinoma	19.2	5637- Bladder carcinoma	6.3
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	76.8
LS 174T- Colon adenocarcinoma	2.9	UM-UC-3- Bladder carcinoma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	4.3	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	18.8	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	1.7	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	19.6	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	1.5	A431- Epidermoid carcinoma	27.9
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	4.7
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	25.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.4	SCC-9- Squamous cell carcinoma of tongue	0.1
RL95-2- Uterine carcinoma	5.3	SCC-15- Squamous cell carcinoma of tongue	1.1
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	37.4

Table ZF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3134, Run 164527747	Tissue Name	Rel. Exp.(%) Ag3134, Run 164527747
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	2.4
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	66.4
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.1
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.2
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	5.6
Primary Th2 act	0.1	Microvascular Dermal EC none	1.3
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	2.6
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	15.0
Primary Th2 rest	0.0	Small airway epithelium none	40.6
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	10.6
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.1
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.1
CD8 lymphocyte act	0.0	Astrocytes rest	2.3
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	1.7
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	25.9
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	100.0
LAK cells IL-2	0.0	Liver cirrhosis	1.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.1
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.4
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	2.7
LAK cells PMA/ionomycin	0.1	NCI-H292 IL-9	0.9
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.4
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	27.9
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.1	Lung fibroblast TNF alpha + IL- 1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.6

Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.1
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	1.8
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	8.8
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.1	Dermal fibroblast CCD1070 TNF alpha	0.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	40.1
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	2.7
Macrophages rest	0.0	Lung	10.8
Macrophages LPS	0.0	Thymus	2.6
HUVEC none	0.0	Kidney	13.8
HUVEC starved	0.0		

- CNS\_neurodegeneration\_v1.0 Summary:** Ag3134 This panel confirms the expression of this gene at low to moderate levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between
- 5 Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.3D for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

- Panel 1.3D Summary:** Ag3134 Expression of the NOV39a gene is highest in placenta (CT = 29.5). In addition, this gene is expressed at moderate levels in all regions of the
- 10 central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord (CTs = 29.5-32). Thus, expression of this gene may be used to distinguish brain and placenta from the other samples on this panel. Furthermore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and
- 15 depression.

- Panel 2D Summary:** Ag3134 Expression of this gene is highest in a bladder cancer sample (CT = 28.3). Interestingly, expression in the matched normal adjacent bladder tissue is much lower (CT = 34.5). In addition, the NOV39a gene is expressed at higher levels in a number of other tumor samples when compared to normal matched adjacent tissue.
- 20 Specifically, expression of this gene is upregulated in gastric cancers and lung cancers. Thus, expression of this gene can be used to distinguish bladder, gastric and lung cancers.

Furthermore, therapeutic modulation of the activity of this gene or its protein product, using small molecule drugs, antibodies, or protein therapeutics, may be of benefit in the treatment of bladder, gastric and lung cancer.

The NOV39a gene encodes a protein with homology to urokinase plasminogen  
5 activator surface receptor precursor, which has previously been shown to play an important role in metastasis of lung and other cancers (Lakka et al., Clin Cancer Res 7(4):1087-93, 2001). In addition, it has been shown that inhibition of urokinase-type plasminogen activator receptor gene using antisense technology reduces tumor cell invasion and metastasis in non-small cell lung cancer cell lines (Lakka et al., Clin Cancer Res 7(4):1087-93, 2001). This observation  
10 suggests that therapeutic inhibition of the NOV39a gene may also be useful for reducing tumor cell invasion and metastasis.

**Panel 3D Summary:** Ag3134 Expression of this gene is highest in a lung cancer cell line (CT = 29). The NOV39a gene is expressed at moderate levels in a number of other cancer cell lines including several lung, gastric, and bladder cancer cell lines. This observation is  
15 consistent with what is seen in Panel 2D.

**Panel 4D Summary:** Ag3134 Expression of the NOV39a gene is upregulated in activated keratinocytes as well as in IFN gamma treated dermal fibroblasts. Therefore, modulation of the activity of the protein encoded by this gene using small molecule drugs or antibodies may be useful in the treatment of psoriasis. The NOV39a gene encodes a protein  
20 with homology to urokinase plasminogen activator surface receptor precursor. Consistent with a potential role for this gene in psoriasis, alterations in plasminogen activator expression have previously been shown to occur in psoriasis (Spiers et al., J Invest Dermatol 102(3):333-8, 1994).

In addition, expression of this gene is upregulated in TNF alpha + IFN gamma treated  
25 HUVEC cells (CT=29.8) and IFN gamma treated NCI-H292 cells (CT=31) as compared to their untreated counterparts (CTs=37-40). This gene also shows a moderate expression in normal lung. The expression of this gene in the activated mucoepidermoid cell line (NCI-H292 cells), and the endothelial cells (HUVEC) suggests that this gene may be important in the proliferation or activation of these cell types. Therefore, therapeutics designed with the  
30 protein encoded by the gene may reduce or eliminate symptoms caused by inflammation in lung epithelia in chronic obstructive pulmonary disease, asthma, allergy, and emphysema.

**AA. NOV40a: novel human agrin**

Expression of gene NOV40a was assessed using the primer-probe set Ag3605, described in Table AAA. Results of the RTQ-PCR runs are shown in Tables AAB, AAC, AAD, AAE and AAF.

Table AAA. Probe Name Ag3605

Primers	Sequences	Length	Start Position
Forward	5'-gacccaagtcagaactgttc-3' (SEQ ID NO:287)	21	3174
Probe	TET-5'-attgagagcacctggacgacctctt-3'-TAMRA (SEQ ID NO:288)	26	3213
Reverse	5'-gaaatccttcttgacgtctgaa-3' (SEQ ID NO:289)	22	3245

5

Table AAB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3605, Run 210997601	Tissue Name	Rel. Exp.(%) Ag3605, Run 210997601
AD 1 Hippo	12.2	Control (Path) 3 Temporal Ctx	10.9
AD 2 Hippo	27.2	Control (Path) 4 Temporal Ctx	57.0
AD 3 Hippo	15.3	AD 1 Occipital Ctx	26.1
AD 4 Hippo	34.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	12.0
AD 6 Hippo	32.3	AD 4 Occipital Ctx	24.7
Control 2 Hippo	37.9	AD 5 Occipital Ctx	44.4
Control 4 Hippo	20.7	AD 6 Occipital Ctx	13.5
Control (Path) 3 Hippo	7.5	Control 1 Occipital Ctx	10.0
AD 1 Temporal Ctx	28.7	Control 2 Occipital Ctx	51.8
AD 2 Temporal Ctx	31.9	Control 3 Occipital Ctx	17.3
AD 3 Temporal Ctx	17.4	Control 4 Occipital Ctx	14.9
AD 4 Temporal Ctx	31.2	Control (Path) 1 Occipital Ctx	90.8
AD 5 Inf Temporal Ctx	87.1	Control (Path) 2 Occipital Ctx	19.8
AD 5 Sup Temporal Ctx	51.4	Control (Path) 3 Occipital Ctx	7.4
AD 6 Inf Temporal Ctx	42.9	Control (Path) 4 Occipital Ctx	44.8
AD 6 Sup Temporal Ctx	51.1	Control 1 Parietal Ctx	14.6
Control 1 Temporal Ctx	17.3	Control 2 Parietal Ctx	53.6
Control 2 Temporal Ctx	50.3	Control 3 Parietal Ctx	18.6
Control 3 Temporal Ctx	23.7	Control (Path) 1 Parietal Ctx	76.3
Control 3 Temporal Ctx	20.3	Control (Path) 2 Parietal Ctx	36.6
Control (Path) 1 Temporal Ctx	78.5	Control (Path) 3 Parietal Ctx	12.0
Control (Path) 2 Temporal Ctx	50.3	Control (Path) 4 Parietal Ctx	64.2



Table AAC. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3605, Run 213406184	Tissue Name	Rel. Exp.(%) Ag3605, Run 213406184
Adipose	1.4	Renal ca. TK-10	19.3
Melanoma* Hs688(A).T	2.6	Bladder	8.4
Melanoma* Hs688(B).T	4.6	Gastric ca. (liver met.) NCI-N87	87.7
Melanoma* M14	6.7	Gastric ca. KATO III	17.8
Melanoma* LOXIMVI	4.8	Colon ca. SW-948	9.2
Melanoma* SK-MEL-5	2.6	Colon ca. SW480	25.0
Squamous cell carcinoma SCC-4	9.0	Colon ca. * (SW480 met) SW620	5.0
Testis Pool	1.3	Colon ca. HT29	26.8
Prostate ca. * (bone met) PC-3	21.0	Colon ca. HCT-116	4.9
Prostate Pool	0.9	Colon ca. CaCo-2	13.6
Placenta	0.9	Colon cancer tissue	10.2
Uterus Pool	0.4	Colon ca. SW1116	5.1
Ovarian ca. OVCAR-3	77.4	Colon ca. Colo-205	1.8
Ovarian ca. SK-OV-3	42.0	Colon ca. SW-48	1.2
Ovarian ca. OVCAR-4	9.5	Colon Pool	1.9
Ovarian ca. OVCAR-5	39.2	Small Intestine Pool	0.6
Ovarian ca. IGROV-1	22.1	Stomach Pool	1.5
Ovarian ca. OVCAR-8	18.0	Bone Marrow Pool	0.6
Ovary	1.5	Fetal Heart	1.4
Breast ca. MCF-7	7.7	Heart Pool	0.7
Breast ca. MDA-MB- 231	22.7	Lymph Node Pool	2.1
Breast ca. BT 549	13.2	Fetal Skeletal Muscle	0.9
Breast ca. T47D	100.0	Skeletal Muscle Pool	0.4
Breast ca. MDA-N	4.8	Spleen Pool	0.7
Breast Pool	1.6	Thymus Pool	1.8
Trachea	2.8	CNS cancer (glio/astro) U87-MG	5.9
Lung	0.2	CNS cancer (glio/astro) U- 118-MG	11.7
Fetal Lung	11.4	CNS cancer (neuro;met) SK-N-AS	1.2
Lung ca. NCI-N417	1.4	CNS cancer (astro) SF-539	6.7
Lung ca. LX-1	10.5	CNS cancer (astro) SNB-75	22.8
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	25.0
Lung ca. SHP-77	1.1	CNS cancer (glio) SF-295	35.1
Lung ca. A549	15.6	Brain (Amygdala) Pool	1.7
Lung ca. NCI-H526	5.4	Brain (cerebellum)	1.4
Lung ca. NCI-H23	18.9	Brain (fetal)	7.0
Lung ca. NCI-H460	11.5	Brain (Hippocampus) Pool	1.6
Lung ca. HOP-62	23.7	Cerebral Cortex Pool	1.9

Lung ca. NCI-H522	1.8	Brain (Substantia nigra) Pool	2.8
Liver	0.5	Brain (Thalamus) Pool	2.7
Fetal Liver	0.8	Brain (whole)	3.4
Liver ca. HepG2	15.5	Spinal Cord Pool	1.8
Kidney Pool	1.5	Adrenal Gland	0.2
Fetal Kidney	5.8	Pituitary gland Pool	0.3
Renal ca. 786-0	46.3	Salivary Gland	1.1
Renal ca. A498	13.8	Thyroid (female)	3.3
Renal ca. ACHN	14.3	Pancreatic ca. CAPAN2	23.7
Renal ca. UO-31	41.5	Pancreas Pool	3.0

Table AAD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3605, Run 173764229	Tissue Name	Rel. Exp.(%) Ag3605, Run 173764229
Normal Colon	4.7	Kidney Margin (OD04348)	100.0
Colon cancer (OD06064)	6.3	Kidney malignant cancer (OD06204B)	21.6
Colon Margin (OD06064)	2.3	Kidney normal adjacent tissue (OD06204E)	23.2
Colon cancer (OD06159)	2.1	Kidney Cancer (OD04450-01)	53.2
Colon Margin (OD06159)	1.8	Kidney Margin (OD04450-03)	21.3
Colon cancer (OD06297-04)	2.0	Kidney Cancer 8120613	1.1
Colon Margin (OD06297-05)	3.0	Kidney Margin 8120614	14.1
CC Gr.2 ascend colon (ODO3921)	3.6	Kidney Cancer 9010320	20.3
CC Margin (ODO3921)	1.3	Kidney Margin 9010321	15.0
Colon cancer metastasis (OD06104)	1.1	Kidney Cancer 8120607	71.7
Lung Margin (OD06104)	1.0	Kidney Margin 8120608	12.2
Colon mets to lung (OD04451-01)	4.5	Normal Uterus	7.3
Lung Margin (OD04451-02)	6.7	Uterine Cancer 064011	6.9
Normal Prostate	2.1	Normal Thyroid	4.3
Prostate Cancer (OD04410)	3.9	Thyroid Cancer 064010	27.0
Prostate Margin (OD04410)	3.4	Thyroid Cancer A302152	19.1
Normal Ovary	7.6	Thyroid Margin A302153	8.1
Ovarian cancer (OD06283-03)	27.9	Normal Breast	14.0
Ovarian Margin (OD06283-07)	2.0	Breast Cancer (OD04566)	13.4
Ovarian Cancer 064008	16.0	Breast Cancer 1024	35.1
Ovarian cancer (OD06145)	10.1	Breast Cancer (OD04590-01)	31.6
Ovarian Margin (OD06145)	8.2	Breast Cancer Mets (OD04590-03)	8.7
Ovarian cancer (OD06455-03)	28.9	Breast Cancer Metastasis (OD04655-05)	13.3

Ovarian Margin (OD06455-07)	1.9	Breast Cancer 064006	21.5
Normal Lung	2.9	Breast Cancer 9100266	17.7
Invasive poor diff. lung adeno (ODO4945-01)	9.4	Breast Margin 9100265	16.5
Lung Margin (ODO4945-03)	7.9	Breast Cancer A209073	13.2
Lung Malignant Cancer (OD03126)	7.5	Breast Margin A2090734	35.4
Lung Margin (OD03126)	7.0	Breast cancer (OD06083)	24.5
Lung Cancer (OD05014A)	17.0	Breast cancer node metastasis (OD06083)	21.5
Lung Margin (OD05014B)	11.7	Normal Liver	5.0
Lung cancer (OD06081)	12.2	Liver Cancer 1026	15.5
Lung Margin (OD06081)	2.4	Liver Cancer 1025	12.2
Lung Cancer (OD04237-01)	1.8	Liver Cancer 6004-T	7.8
Lung Margin (OD04237-02)	16.2	Liver Tissue 6004-N	6.1
Ocular Melanoma Metastasis	8.4	Liver Cancer 6005-T	25.0
Ocular Melanoma Margin (Liver)	2.9	Liver Tissue 6005-N	12.4
Melanoma Metastasis	4.0	Liver Cancer 064003	12.9
Melanoma Margin (Lung)	3.8	Normal Bladder	14.2
Normal Kidney	10.9	Bladder Cancer 1023	9.5
Kidney Ca, Nuclear grade 2 (OD04338)	35.4	Bladder Cancer A302173	12.2
Kidney Margin (OD04338)	20.4	Normal Stomach	8.7
Kidney Ca Nuclear grade 1/2 (OD04339)	52.9	Gastric Cancer 9060397	8.9
Kidney Margin (OD04339)	16.6	Stomach Margin 9060396	7.4
Kidney Ca, Clear cell type (OD04340)	16.6	Gastric Cancer 9060395	7.0
Kidney Margin (OD04340)	7.4	Stomach Margin 9060394	7.5
Kidney Ca, Nuclear grade 3 (OD04348)	11.2	Gastric Cancer 064005	6.9

Table AAE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3605, Run 169943454	Tissue Name	Rel. Exp.(%) Ag3605, Run 169943454
Secondary Th1 act	1.0	HUVEC IL-1beta	15.6
Secondary Th2 act	5.1	HUVEC IFN gamma	12.9
Secondary Tr1 act	2.5	HUVEC TNF alpha + IFN gamma	37.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	31.4
Secondary Th2 rest	0.4	HUVEC IL-11	14.9
Secondary Tr1 rest	0.4	Lung Microvascular EC none	79.0
Primary Th1 act	3.6	Lung Microvascular EC TNFalpha + IL-1beta	100.0
Primary Th2 act	1.1	Microvascular Dermal EC none	49.7
Primary Tr1 act	3.4	Microvascular Dermal EC TNFalpha + IL-1beta	56.6

Primary Th1 rest	0.9	Bronchial epithelium TNFalpha + IL1beta	78.5
Primary Th2 rest	0.5	Small airway epithelium none	31.0
Primary Tr1 rest	0.2	Small airway epithelium TNFalpha + IL-1beta	81.8
CD45RA CD4 lymphocyte act	43.5	Coronary artery SMC rest	17.2
CD45RO CD4 lymphocyte act	5.0	Coronary artery SMC TNFalpha + IL-1beta	22.2
CD8 lymphocyte act	3.9	Astrocytes rest	80.1
Secondary CD8 lymphocyte rest	3.7	Astrocytes TNFalpha + IL-1beta	82.9
Secondary CD8 lymphocyte act	3.3	KU-812 (Basophil) rest	2.6
CD4 lymphocyte none	0.3	KU-812 (Basophil) PMA/ionomycin	0.6
2ry Th1/Th2/Tr1 anti-CD95 CH11	0.3	CCD1106 (Keratinocytes) none	70.7
LAK cells rest	5.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	77.4
LAK cells IL-2	2.8	Liver cirrhosis	13.2
LAK cells IL-2+IL-12	1.4	NCI-H292 none	57.8
LAK cells IL-2+IFN gamma	2.3	NCI-H292 IL-4	62.4
LAK cells IL-2+ IL-18	2.9	NCI-H292 IL-9	61.6
LAK cells PMA/ionomycin	6.8	NCI-H292 IL-13	53.6
NK Cells IL-2 rest	1.6	NCI-H292 IFN gamma	67.4
Two Way MLR 3 day	10.7	HPAEC none	21.5
Two Way MLR 5 day	6.5	HPAEC TNF alpha + IL-1 beta	37.6
Two Way MLR 7 day	4.3	Lung fibroblast none	22.4
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	71.2
PBMC PWM	5.0	Lung fibroblast IL-4	16.2
PBMC PHA-L	5.5	Lung fibroblast IL-9	31.9
Ramos (B cell) none	0.4	Lung fibroblast IL-13	18.7
Ramos (B cell) ionomycin	0.2	Lung fibroblast IFN gamma	23.0
B lymphocytes PWM	2.4	Dermal fibroblast CCD1070 rest	15.9
B lymphocytes CD40L and IL-4	1.5	Dermal fibroblast CCD1070 TNF alpha	15.9
EOL-1 dbcAMP	3.4	Dermal fibroblast CCD1070 IL-1 beta	17.2
EOL-1 dbcAMP PMA/ionomycin	18.3	Dermal fibroblast IFN gamma	7.2
Dendritic cells none	14.8	Dermal fibroblast IL-4	7.2
Dendritic cells LPS	48.3	Dermal Fibroblasts rest	4.3
Dendritic cells anti-CD40	9.7	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.9	Neutrophils rest	0.2
Monocytes LPS	66.4	Colon	7.0
Macrophages rest	16.2	Lung	23.3
Macrophages LPS	54.7	Thymus	5.8

HUVEC none	9.3	Kidney	23.2
HUVEC starved	14.4		

Table AAF. Panel CNS\_1

Tissue Name	Rel. Exp.(%) Ag3605, Run 171648697	Tissue Name	Rel. Exp.(%) Ag3605, Run 171648697
BA4 Control	23.8	BA17 PSP	21.9
BA4 Control2	44.4	BA17 PSP2	14.4
BA4 Alzheimer's2	7.5	Sub Nigra Control	37.9
BA4 Parkinson's	66.0	Sub Nigra Control2	31.0
BA4 Parkinson's2	80.7	Sub Nigra Alzheimer's2	26.4
BA4 Huntington's	23.5	Sub Nigra Parkinson's2	80.1
BA4 Huntington's2	49.3	Sub Nigra Huntington's	76.3
BA4 PSP	19.5	Sub Nigra Huntington's2	29.7
BA4 PSP2	34.9	Sub Nigra PSP2	11.1
BA4 Depression	20.6	Sub Nigra Depression	34.4
BA4 Depression2	21.3	Sub Nigra Depression2	18.8
BA7 Control	53.2	Glob Palladus Control	40.3
BA7 Control2	47.6	Glob Palladus Control2	35.8
BA7 Alzheimer's2	13.6	Glob Palladus Alzheimer's	20.0
BA7 Parkinson's	39.8	Glob Palladus Alzheimer's2	21.8
BA7 Parkinson's2	60.7	Glob Palladus Parkinson's	100.0
BA7 Huntington's	41.8	Glob Palladus Parkinson's2	25.0
BA7 Huntington's2	62.9	Glob Palladus PSP	17.9
BA7 PSP	36.1	Glob Palladus PSP2	7.2
BA7 PSP2	25.0	Glob Palladus Depression	15.5
BA7 Depression	20.0	Temp Pole Control	15.3
BA9 Control	36.6	Temp Pole Control2	76.8
BA9 Control2	83.5	Temp Pole Alzheimer's	14.3
BA9 Alzheimer's	17.1	Temp Pole Alzheimer's2	14.7
BA9 Alzheimer's2	34.4	Temp Pole Parkinson's	76.3
BA9 Parkinson's	70.7	Temp Pole Parkinson's2	77.4
BA9 Parkinson's2	74.2	Temp Pole Huntington's	39.8
BA9 Huntington's	55.5	Temp Pole PSP	7.4
BA9 Huntington's2	45.1	Temp Pole PSP2	8.1
BA9 PSP	28.7	Temp Pole Depression2	31.6
BA9 PSP2	8.2	Cing Gyr Control	82.4
BA9 Depression	18.0	Cing Gyr Control2	82.4
BA9 Depression2	0.0	Cing Gyr Alzheimer's	27.4
BA17 Control	74.7	Cing Gyr Alzheimer's2	36.3

BA17 Control2	86.5	Cing Gyr Parkinson's	46.3
BA17 Alzheimer's2	20.3	Cing Gyr Parkinson's2	42.6
BA17 Parkinson's	75.3	Cing Gyr Huntington's	70.7
BA17 Parkinson's2	85.3	Cing Gyr Huntington's2	37.6
BA17 Huntington's	47.0	Cing Gyr PSP	21.6
BA17 Huntington's2	26.6	Cing Gyr PSP2	13.9
BA17 Depression	24.8	Cing Gyr Depression	21.3
BA17 Depression2	41.2	Cing Gyr Depression2	32.5

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3605 This panel confirms the expression of this gene at moderate levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased

5 postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

**General\_screening\_panel\_v1.4 Summary:** Ag3605 Expression of the NOV40a gene is highest in a breast cancer cell line (CT = 25.2). In addition, expression of this gene is

10 primarily associated with cancer cell lines rather than with normal tissues. Specifically, expression of this gene is upregulated in pancreatic, CNS, colon, gastric, renal, lung, breast, ovarian, and prostate cancer cell lines when compared to their respective normal tissues. Thus, therapeutic modulation of the activity of this gene or its protein product, using small molecule drugs, antibodies or protein therapeutics, may be of benefit in the treatment of these types of

15 cancers.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. The NOV40a gene encodes a protein with homology to agrin, a neuronal aggregating factor that induces the aggregation of acetylcholine

20 receptors and other postsynaptic proteins on muscle fibers and is crucial for the formation of the neuromuscular junction. More recently, it has been shown that agrin plays an important role in defining neuronal responses to excitatory neurotransmitters both in vitro and in vivo (Hilgenberg et al., Mol Cell Neurosci 19(1):97-110, 2002; Bixby et al., J Neurobiol 50(2):164-79, 2002). The NOV40a gene expression in the central nervous system is consistent with the

25 hypothesis that this protein may have similar functions as agrin. Therefore, this gene may play

a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, thyroid, and the gastrointestinal tract and at low levels in adrenal gland, pituitary gland, skeletal muscle, heart, and liver. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes. In support of this hypothesis, decreased glomerular expression of agrin in has been observed in diabetic nephropathy (Yard et al., Decreased glomerular expression of agrin in diabetic nephropathy and podocytes, cultured in high glucose medium. Exp Nephrol 9(3):214-22, 2001).

**Panel 2.2 Summary:** Ag3605 Expression of the NOV40a gene is highest in a sample of normal kidney (CT = 27.4). Interestingly, expression of this gene appears to be upregulated in a number of ovarian and renal cancers when compared to the matched control margins. Thus, expression of this gene could be used as a marker for ovarian and renal carcinoma. Furthermore, therapeutic modulation of the activity of this gene or its protein product, using small molecule drugs, antibodies or protein therapeutics, could be of benefit in the treatment of renal and ovarian cancer. This gene is expressed at moderate levels in the remaining samples on this panel, with little or no difference in expression levels between tumor and normal tissue.

**Panel 4.1D Summary:** Ag3605 Expression of the NOV40a gene is highest in lung microvascular endothelial cells, microvascular dermal endothelial cells, mucoepidermoid cell line NCI-H292, astrocytes, and keratinocytes. Therefore, small molecule drug, antibody or protein therapeutics designed against the protein encoded by the NOV40a gene could reduce or inhibit inflammation in asthma, emphysema, allergy, psoriasis, muscular dystrophy and multiple sclerosis.

The NOV40a gene encodes a protein with homology to agrin. Recently, it has been demonstrated that agrin, an aggregating protein crucial for formation of the neuromuscular junction, is also expressed in lymphocytes and is important in reorganization of membrane lipid microdomains and setting the threshold for T cell signaling (Khan et al., Science 292(5522):1681-6, 2001). T cell activation is dependent on both a primary signal delivered through the T cell receptor and a secondary costimulatory signal mediated by coreceptors. Costimulation is thought to act through the specific redistribution and clustering of membrane and intracellular kinase-rich lipid raft microdomains at the contact site between T cells and antigen-presenting cells. This site has been termed the immunologic synapse. Khan et al.

(2001) concluded that agrin induces the aggregation of signaling proteins and the creation of signaling domains in both immune and nervous systems through a common lipid raft pathway.

**Panel CNS\_1 Summary:** Ag3605 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

#### AB. NOV41a: MAJOR URINARY PROTEIN 4 PRECURSOR (MUP 4)

Expression of gene NOV41a was assessed using the primer-probe set Ag2289, described in Table ABA. Results of the RTQ-PCR runs are shown in Tables ABB, ABC, and ABD.

Table ABA. Probe Name Ag2289

Primers	Sequences	Length	Start Position
Forward	5'-gagccactgctagagaagac-3' (SEQ ID NO:290)	22	55
Probe	TET-5'-tgctgtcccttaccagatgatgctg-3'-TAMRA (SEQ ID NO:291)	26	105
Reverse	5'-acccagacacagcaacag-3' (SEQ ID NO:292)	19	131

Table ABB. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag2289, Run 209731955	Tissue Name	Rel. Exp.(%) Ag2289, Run 209731955
AD 1 Hippo	1.0	Control (Path) 3 Temporal Ctx	0.5
AD 2 Hippo	14.6	Control (Path) 4 Temporal Ctx	14.9
AD 3 Hippo	0.0	AD 1 Occipital Ctx	10.7
AD 4 Hippo	6.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	29.7	AD 3 Occipital Ctx	0.0
AD 6 Hippo	100.0	AD 4 Occipital Ctx	6.7
Control 2 Hippo	4.2	AD 5 Occipital Ctx	5.6
Control 4 Hippo	2.1	AD 6 Occipital Ctx	16.2
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	1.4
AD 1 Temporal Ctx	7.1	Control 2 Occipital Ctx	27.4
AD 2 Temporal Ctx	5.9	Control 3 Occipital Ctx	7.6
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	4.3
AD 4 Temporal Ctx	9.8	Control (Path) 1 Occipital Ctx	28.9
AD 5 Inf Temporal Ctx	8.4	Control (Path) 2 Occipital Ctx	7.0
AD 5 SupTemporal Ctx	1.9	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	47.3	Control (Path) 4 Occipital Ctx	2.1
AD 6 Sup Temporal Ctx	61.1	Control 1 Parietal Ctx	1.8



Control 1 Temporal Ctx	2.2	Control 2 Parietal Ctx	13.1
Control 2 Temporal Ctx	29.7	Control 3 Parietal Ctx	11.0
Control 3 Temporal Ctx	6.9	Control (Path) 1 Parietal Ctx	29.7
Control 4 Temporal Ctx	0.0	Control (Path) 2 Parietal Ctx	6.9
Control (Path) 1 Temporal Ctx	4.9	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	16.8	Control (Path) 4 Parietal Ctx	1.2

Table ABC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2289, Run 151630364	Tissue Name	Rel. Exp.(%) Ag2289, Run 151630364
Liver adenocarcinoma	4.8	Kidney (fetal)	3.3
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	6.1
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	86.5	Liver	0.0
Brain (whole)	38.2	Liver (fetal)	0.0
Brain (amygdala)	51.8	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	3.6	Lung	6.1
Brain (hippocampus)	100.0	Lung (fetal)	4.2
Brain (substantia nigra)	18.8	Lung ca. (small cell) LX- 1	22.5
Brain (thalamus)	20.9	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	66.9	Lung ca. (s.cell var.) SHP-77	4.3
Spinal cord	0.0	Lung ca. (large cell) NCI- H460	0.0
glio/astro U87-MG	3.3	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	2.6	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI- H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI- H596	0.0
glioma SNB-19	3.4	Mammary gland	4.2
glioma U251	0.0	Breast ca.* (pl.ef) MCF- 7	0.0

glioma SF-295	5.7	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	25.3	Breast ca. MDA-N	6.2
Skeletal muscle	0.0	Ovary	5.4
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	3.2	Ovarian ca. OVCAR-5	0.0
Lymph node	3.9	Ovarian ca. OVCAR-8	25.9
Colorectal	16.4	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.3	Placenta	0.0
Colon ca.* SW620(SW480 met)	5.1	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	3.4	Testis	2.5
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	6.0	Adipose	0.0

Table ABD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag2289, Run 169828803	Tissue Name	Rel. Exp.(%) Ag2289, Run 169828803
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	5.7	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0

Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	14.4	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	12.3	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	15.1
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	27.9
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	27.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	18.6	NCI-H292 IFN gamma	12.7
Two Way MLR 3 day	59.5	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL- 1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	22.1	Dermal fibroblast CCD1070 IL- 1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	12.9	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	26.2	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	42.6	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	19.9	Lung	18.4
Macrophages LPS	0.0	Thymus	15.3
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

**CNS\_neurodegeneration\_v1.0 Summary:** Ag2289 This panel does not show differential expression of the NOV41a gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

- 5        **Panel 1.3D Summary:** Ag2289 Expression of the NOV41a gene appears to be highly brain specific, with highest expression in hippocampus (CT=30). Therefore, this gene may play a role in central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

10        In addition, this gene is expressed at much higher levels in fetal skeletal muscle tissue (CT=32) when compared to expression in the adult counterpart (CT=40). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue. Furthermore, the relative overexpression of this gene in fetal skeletal muscle suggests that the protein product may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein  
15        encoded by this gene could be useful in treatment of muscle related diseases. More specifically, treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function.

20        This gene is a homolog of MUP, whose murine homolog has been shown to have pheromone binding activity (Timm et al., Protein Sci 10(5):997-1004, 2001; Novotny et al., Proc R Soc Lond B Biol Sci 266(1432):2017-22, 1999). Based on the homology, this protein may play a role in sexual maturation and cycling in adult females.

**Panel 2.2 Summary:** Ag2289 Expression of the NOV41a gene is low/undetectable in all samples on this panel (CTs>35).

25        **Panel 4.1D Summary:** Ag2289 The NOV41a gene is only expressed at detectable levels in the kidney (CT=34.7). The putative protein encoded for by this gene may allow cells within the kidney to respond to specific microenvironmental signals. Therefore, therapies designed with the protein encoded by this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

30        **Panel CNS\_1 Summary:** Ag2289 Expression of the NOV41a gene is low/undetectable in all samples on this panel (CTs>35).

#### **AC. NOV41b: MAJOR URINARY PROTEIN 4 PRECURSOR**

Expression of gene NOV41b was assessed using the primer-probe set Ag2321, described in Table ACA.

Table ACA. Probe Name Ag2321

Primers	Sequences	Length	Start Position
Forward	5'-caggaggaagaaaacaatgatg-3' (SEQ ID NO:293)	22	73
Probe	TET-5'-tgtgacaagcaacttcgatctgtcaa-3'-TAMRA (SEQ ID NO:294)	26	96
Reverse	5'-aaccgaataccactctcctgaa-3' (SEQ ID NO:295)	22	126

**Panel 1.3D Summary:** Ag2321 Expression of the NOV41b gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 2.2 Summary:** Ag2321 Expression of the NOV41b gene is low/undetectable in all samples on this panel (CTs>35).

**Panel 4D Summary:** Ag2321 Expression of the NOV41b gene is low/undetectable in all samples on this panel (CTs>35).

10

#### **AD. NOV42a and CG59889-02 and NOV42c: KIAA1199**

Expression of gene NOV42a, variant CG59889-02 and full length clone NOV42c was assessed using the primer-probe set Ag3626, described in Table ADA. Results of the RTQ-PCR runs are shown in Tables ADB, ADC, ADD, ADE and ADF. Please note that NOV42c represents a full-length physical clone of the NOV42c gene, validating the prediction of the gene sequence

15

Table ADA. Probe Name Ag3626

Primers	Sequences	Length	Start Position
Forward	5'-ctgaggatcacaagccaaa-3' (SEQ ID NO:296)	20	4091
Probe	TET-5'-atcttccaagtgtgcccacccctgt-3'-TAMRA (SEQ ID NO:297)	26	4111
Reverse	5'-cagctgtcctcacaacttcttc-3' (SEQ ID NO:298)	22	4146

Table ADB. AI\_comprehensive panel\_v1.0

Tissue Name	Rel. Exp.(%) Ag3626, Run 234222205	Tissue Name	Rel. Exp.(%) Ag3626, Run 234222205
110967 COPD-F	1.0	112427 Match Control Psoriasis-F	13.7
110980 COPD-F	1.6	112418 Psoriasis-M	1.8
110968 COPD-M	2.2	112723 Match Control Psoriasis-M	15.1
110977 COPD-M	8.5	112419 Psoriasis-M	4.6
110989 Emphysema-F	16.3	112424 Match Control Psoriasis-M	2.0
110992 Emphysema-F	4.3	112420 Psoriasis-M	12.2

110993 Emphysema-F	3.3	112425 Match Control Psoriasis-M	9.6
110994 Emphysema-F	1.2	104689 (MF) OA Bone-Backus	22.7
110995 Emphysema-F	11.6	104690 (MF) Adj "Normal" Bone-Backus	12.4
110996 Emphysema-F	1.6	104691 (MF) OA Synovium-Backus	28.5
110997 Asthma-M	0.9	104692 (BA) OA Cartilage-Backus	45.1
111001 Asthma-F	2.6	104694 (BA) OA Bone-Backus	39.8
111002 Asthma-F	9.2	104695 (BA) Adj "Normal" Bone-Backus	26.2
111003 Atopic Asthma-F	4.0	104696 (BA) OA Synovium-Backus	45.4
111004 Atopic Asthma-F	7.6	104700 (SS) OA Bone-Backus	13.1
111005 Atopic Asthma-F	2.0	104701 (SS) Adj "Normal" Bone-Backus	31.6
111006 Atopic Asthma-F	2.4	104702 (SS) OA Synovium-Backus	13.0
111417 Allergy-M	3.4	117093 OA Cartilage Rep7	8.4
112347 Allergy-M	0.4	112672 OA Bone5	31.6
112349 Normal Lung-F	0.1	112673 OA Synovium5	15.7
112357 Normal Lung-F	13.8	112674 OA Synovial Fluid cells5	15.1
112354 Normal Lung-M	1.5	117100 OA Cartilage Rep14	2.3
112374 Crohns-F	28.9	112756 OA Bone9	100.0
112389 Match Control Crohns-F	3.5	112757 OA Synovium9	0.6
112375 Crohns-F	43.8	112758 OA Synovial Fluid Cells9	1.9
112732 Match Control Crohns-F	8.2	117125 RA Cartilage Rep2	1.3
112725 Crohns-M	6.1	113492 Bone2 RA	4.2
112387 Match Control Crohns-M	15.6	113493 Synovium2 RA	2.0
112378 Crohns-M	0.2	113494 Syn Fluid Cells RA	5.5
112390 Match Control Crohns-M	16.8	113499 Cartilage4 RA	4.3
112726 Crohns-M	6.5	113500 Bone4 RA	9.5
112731 Match Control Crohns-M	6.1	113501 Synovium4 RA	6.1
112380 Ulcer Col-F	5.0	113502 Syn Fluid Cells4 RA	3.7
112734 Match Control Ulcer Col-F	29.9	113495 Cartilage3 RA	3.9
112384 Ulcer Col-F	21.9	113496 Bone3 RA	7.4
112737 Match Control Ulcer Col-F	0.5	113497 Synovium3 RA	2.4
112386 Ulcer Col-F	0.9	113498 Syn Fluid Cells3 RA	3.3
112738 Match Control	2.0	117106 Normal Cartilage	1.8

Ulcer Col-F		Rep20	
112381 Ulcer Col-M	0.1	113663 Bone3 Normal	0.8
112735 Match Control Ulcer Col-M	8.5	113664 Synovium3 Normal	0.1
112382 Ulcer Col-M	4.0	113665 Syn Fluid Cells3 Normal	0.2
112394 Match Control Ulcer Col-M	2.0	117107 Normal Cartilage Rep22	1.2
112383 Ulcer Col-M	14.9	113667 Bone4 Normal	8.1
112736 Match Control Ulcer Col-M	4.4	113668 Synovium4 Normal	6.0
112423 Psoriasis-F	3.3	113669 Syn Fluid Cells4 Normal	17.0

Table ADC. CNS\_neurodegeneration\_v1.0

Tissue Name	Rel. Exp.(%) Ag3626, Run 206916253	Tissue Name	Rel. Exp.(%) Ag3626, Run 206916253
AD 1 Hippo	17.7	Control (Path) 3 Temporal Ctx	8.2
AD 2 Hippo	26.6	Control (Path) 4 Temporal Ctx	23.7
AD 3 Hippo	10.1	AD 1 Occipital Ctx	25.5
AD 4 Hippo	6.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	11.4
AD 6 Hippo	31.4	AD 4 Occipital Ctx	12.2
Control 2 Hippo	10.9	AD 5 Occipital Ctx	55.9
Control 4 Hippo	31.4	AD 6 Occipital Ctx	35.8
Control (Path) 3 Hippo	14.6	Control 1 Occipital Ctx	22.8
AD 1 Temporal Ctx	25.7	Control 2 Occipital Ctx	69.3
AD 2 Temporal Ctx	33.0	Control 3 Occipital Ctx	23.7
AD 3 Temporal Ctx	12.9	Control 4 Occipital Ctx	9.7
AD 4 Temporal Ctx	16.5	Control (Path) 1 Occipital Ctx	63.3
AD 5 Inf Temporal Ctx	72.7	Control (Path) 2 Occipital Ctx	19.8
AD 5 Sup Temporal Ctx	46.3	Control (Path) 3 Occipital Ctx	1.9
AD 6 Inf Temporal Ctx	38.4	Control (Path) 4 Occipital Ctx	56.3
AD 6 Sup Temporal Ctx	43.8	Control 1 Parietal Ctx	13.2
Control 1 Temporal Ctx	22.2	Control 2 Parietal Ctx	47.6
Control 2 Temporal Ctx	20.3	Control 3 Parietal Ctx	12.1
Control 3 Temporal Ctx	12.6	Control (Path) 1 Parietal Ctx	66.9
Control 4 Temporal Ctx	10.7	Control (Path) 2 Parietal Ctx	46.3
Control (Path) 1 Temporal Ctx	45.1	Control (Path) 3 Parietal Ctx	6.2
Control (Path) 2	21.2	Control (Path) 4	51.4

Temporal Ctx		Parietal Ctx	
--------------	--	--------------	--

Table ADD. General\_screening\_panel\_v1.4

Tissue Name	Rel. Exp.(%) Ag3626, Run 213406211	Tissue Name	Rel. Exp.(%) Ag3626, Run 213406211
Adipose	0.0	Renal ca. TK-10	2.9
Melanoma* Hs688(A).T	61.1	Bladder	0.7
Melanoma* Hs688(B).T	86.5	Gastric ca. (liver met.) NCI-N87	1.8
Melanoma* M14	0.5	Gastric ca. KATO III	100.0
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	9.9
Melanoma* SK-MEL-5	1.6	Colon ca. SW480	0.9
Squamous cell carcinoma SCC-4	0.3	Colon ca.* (SW480 met) SW620	19.2
Testis Pool	0.3	Colon ca. HT29	3.8
Prostate ca.* (bone met) PC-3	0.6	Colon ca. HCT-116	0.5
Prostate Pool	0.1	Colon ca. CaCo-2	0.1
Placenta	0.4	Colon cancer tissue	9.5
Uterus Pool	0.0	Colon ca. SW1116	2.5
Ovarian ca. OVCAR-3	0.5	Colon ca. Colo-205	6.0
Ovarian ca. SK-OV-3	0.2	Colon ca. SW-48	6.3
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.1
Ovarian ca. OVCAR-5	1.4	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.2	Stomach Pool	0.5
Ovarian ca. OVCAR-8	0.7	Bone Marrow Pool	0.1
Ovary	0.6	Fetal Heart	0.1
Breast ca. MCF-7	0.1	Heart Pool	0.0
Breast ca. MDA-MB-231	25.2	Lymph Node Pool	0.1
Breast ca. BT 549	0.1	Fetal Skeletal Muscle	0.0
Breast ca. T47D	4.8	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.4	Spleen Pool	0.1
Breast Pool	0.2	Thymus Pool	0.5
Trachea	0.6	CNS cancer (glio/astro) U87-MG	0.6
Lung	0.1	CNS cancer (glio/astro) U-118-MG	2.0
Fetal Lung	0.5	CNS cancer (neuro;met) SK-N-AS	0.4
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	1.9
Lung ca. LX-1	36.3	CNS cancer (astro) SNB-75	3.3
Lung ca. NCI-H146	4.2	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	2.1	CNS cancer (glio) SF-295	0.3
Lung ca. A549	0.2	Brain (Amygdala) Pool	0.2
Lung ca. NCI-H526	0.0	Brain (cerebellum)	1.2
Lung ca. NCI-H23	0.7	Brain (fetal)	0.7
Lung ca. NCI-H460	0.2	Brain (Hippocampus) Pool	0.4



Lung ca. HOP-62	0.1	Cerebral Cortex Pool	0.5
Lung ca. NCI-H522	0.1	Brain (Substantia nigra) Pool	0.3
Liver	0.0	Brain (Thalamus) Pool	0.4
Fetal Liver	0.0	Brain (whole)	0.9
Liver ca. HepG2	8.7	Spinal Cord Pool	1.3
Kidney Pool	0.1	Adrenal Gland	0.1
Fetal Kidney	0.1	Pituitary gland Pool	0.1
Renal ca. 786-0	0.1	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.1
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.1
Renal ca. UO-31	0.1	Pancreas Pool	0.2

Table ADE. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3626, Run 173764230	Tissue Name	Rel. Exp.(%) Ag3626, Run 173764230
Normal Colon	13.8	Kidney Margin (OD04348)	16.6
Colon cancer (OD06064)	43.5	Kidney malignant cancer (OD06204B)	7.5
Colon Margin (OD06064)	5.8	Kidney normal adjacent tissue (OD06204E)	3.7
Colon cancer (OD06159)	5.3	Kidney Cancer (OD04450-01)	12.5
Colon Margin (OD06159)	4.0	Kidney Margin (OD04450-03)	4.0
Colon cancer (OD06297-04)	52.5	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	9.2	Kidney Margin 8120614	10.3
CC Gr.2 ascend colon (ODO3921)	47.6	Kidney Cancer 9010320	20.2
CC Margin (ODO3921)	8.0	Kidney Margin 9010321	7.3
Colon cancer metastasis (OD06104)	15.7	Kidney Cancer 8120607	14.3
Lung Margin (OD06104)	0.0	Kidney Margin 8120608	4.8
Colon mets to lung (OD04451-01)	83.5	Normal Uterus	5.3
Lung Margin (OD04451-02)	4.0	Uterine Cancer 064011	8.5
Normal Prostate	0.0	Normal Thyroid	0.0
Prostate Cancer (OD04410)	4.0	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	4.5	Thyroid Cancer A302152	30.8
Normal Ovary	21.9	Thyroid Margin A302153	0.0
Ovarian cancer (OD06283-03)	58.2	Normal Breast	9.5
Ovarian Margin (OD06283-07)	22.1	Breast Cancer (OD04566)	4.6
Ovarian Cancer 064008	42.9	Breast Cancer 1024	11.6
Ovarian cancer (OD06145)	33.4	Breast Cancer (OD04590-01)	34.2
Ovarian Margin (OD06145)	51.4	Breast Cancer Mets (OD04590-03)	14.1

Ovarian cancer (OD06455-03)	26.6	Breast Cancer Metastasis (OD04655-05)	36.6
Ovarian Margin (OD06455-07)	0.9	Breast Cancer 064006	14.9
Normal Lung	10.2	Breast Cancer 9100266	2.9
Invasive poor diff. lung adeno (ODO4945-01)	8.7	Breast Margin 9100265	9.8
Lung Margin (ODO4945-03)	17.8	Breast Cancer A209073	7.6
Lung Malignant Cancer (OD03126)	45.1	Breast Margin A2090734	9.4
Lung Margin (OD03126)	10.7	Breast cancer (OD06083)	17.4
Lung Cancer (OD05014A)	60.3	Breast cancer node metastasis (OD06083)	33.9
Lung Margin (OD05014B)	9.5	Normal Liver	5.5
Lung cancer (OD06081)	18.2	Liver Cancer 1026	18.3
Lung Margin (OD06081)	5.4	Liver Cancer 1025	5.0
Lung Cancer (OD04237-01)	12.7	Liver Cancer 6004-T	18.3
Lung Margin (OD04237-02)	100.0	Liver Tissue 6004-N	9.7
Ocular Melanoma Metastasis	4.6	Liver Cancer 6005-T	30.1
Ocular Melanoma Margin (Liver)	4.0	Liver Tissue 6005-N	14.7
Melanoma Metastasis	4.7	Liver Cancer 064003	8.1
Melanoma Margin (Lung)	1.9	Normal Bladder	10.2
Normal Kidney	3.0	Bladder Cancer 1023	39.2
Kidney Ca, Nuclear grade 2 (OD04338)	4.6	Bladder Cancer A302173	41.8
Kidney Margin (OD04338)	5.2	Normal Stomach	6.3
Kidney Ca Nuclear grade 1/2 (OD04339)	18.2	Gastric Cancer 9060397	21.9
Kidney Margin (OD04339)	5.0	Stomach Margin 9060396	20.2
Kidney Ca, Clear cell type (OD04340)	2.0	Gastric Cancer 9060395	27.5
Kidney Margin (OD04340)	11.7	Stomach Margin 9060394	4.7
Kidney Ca, Nuclear grade 3 (OD04348)	3.9	Gastric Cancer 064005	17.6

Table ADF. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3626, Run 182098824	Tissue Name	Rel. Exp.(%) Ag3626, Run 182098824
Daoy- Medulloblastoma	0.2	Ca Ski- Cervical epidermoid carcinoma (metastasis)	15.2
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.4
D283 Med- Medulloblastoma	2.2	Ramos- Stimulated with PMA/ionomycin 6h	0.2
PFSK-1- Primitive Neuroectodermal	1.2	Ramos- Stimulated with PMA/ionomycin 14h	0.4
XF-498- CNS	0.6	MEG-01- Chronic myelogenous leukemia (megakaryoblast)	0.6
SNB-78- Glioma	11.0	Raji- Burkitt's lymphoma	0.3

SF-268- Glioblastoma	0.6	Daudi- Burkitt's lymphoma	0.5
T98G- Glioblastoma	4.1	U266- B-cell plasmacytoma	0.2
SK-N-SH- Neuroblastoma (metastasis)	1.1	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.5	RL- non-Hodgkin's B-cell lymphoma	0.3
Cerebellum	1.4	JM1- pre-B-cell lymphoma	2.1
Cerebellum	4.2	Jurkat- T cell leukemia	0.5
NCI-H292- Mucoepidermoid lung carcinoma	5.8	TF-1- Erythroleukemia	0.5
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	1.0
DMS-79- Small cell lung cancer	16.2	U937- Histiocytic lymphoma	0.1
NCI-H146- Small cell lung cancer	4.9	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	1.1	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.2	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	2.1	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	19.3	Hs766T- Pancreatic carcinoma (LN metastasis)	28.7
NCI-H1299- Large cell lung cancer	0.8	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.4
NCI-H727- Lung carcinoid	17.0	SU86.86- Pancreatic carcinoma (liver metastasis)	1.1
NCI-UMC-11- Lung carcinoid	1.9	BxPC-3- Pancreatic adenocarcinoma	4.1
LX-1- Small cell lung cancer	100.0	HPAC- Pancreatic adenocarcinoma	7.2
Colo-205- Colon cancer	31.2	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	57.8	CFPAC-1- Pancreatic ductal adenocarcinoma	2.2
KM20L2- Colon cancer	14.4	PANC-1- Pancreatic epithelioid ductal carcinoma	0.4
NCI-H716- Colon cancer	1.6	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	26.8	5637- Bladder carcinoma	0.2
SW1116- Colon adenocarcinoma	9.9	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	17.3	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	4.2	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	21.0	HT-1080- Fibrosarcoma	30.6
NCI-SNU-5- Gastric carcinoma	0.5	MG-63- Osteosarcoma	0.3

KATO III- Gastric carcinoma	0.9	SK-LMS-1- Leiomyosarcoma (vulva)	0.2
NCI-SNU-16- Gastric carcinoma	0.7	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.3
NCI-SNU-1- Gastric carcinoma	0.3	A431- Epidermoid carcinoma	0.1
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	10.4
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.8
MKN-45- Gastric carcinoma	32.3	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	29.3	SCC-4- Squamous cell carcinoma of tongue	0.1
OVCAR-5- Ovarian carcinoma	0.8	SCC-9- Squamous cell carcinoma of tongue	0.2
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.2
HelaS3- Cervical adenocarcinoma	0.3	CAL 27- Squamous cell carcinoma of tongue	0.5

Table ADG. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3626, Run 169946026	Tissue Name	Rel. Exp.(%) Ag3626, Run 169946026
Secondary Th1 act	0.4	HUVEC IL-1 beta	0.2
Secondary Th2 act	0.1	HUVEC IFN gamma	0.2
Secondary Tr1 act	0.3	HUVEC TNF alpha + IFN gamma	0.2
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.1
Secondary Th2 rest	0.6	HUVEC IL-11	0.2
Secondary Tr1 rest	0.2	Lung Microvascular EC none	0.7
Primary Th1 act	0.3	Lung Microvascular EC TNFalpha + IL-1 beta	1.2
Primary Th2 act	0.6	Microvascular Dermal EC none	0.1
Primary Tr1 act	0.6	Microvascular Dermal EC TNFalpha + IL-1 beta	0.7
Primary Th1 rest	0.2	Bronchial epithelium TNFalpha + IL1 beta	0.5
Primary Th2 rest	0.2	Small airway epithelium none	1.1
Primary Tr1 rest	0.3	Small airway epithelium TNFalpha + IL-1 beta	1.1
CD45RA CD4 lymphocyte act	29.1	Coronary artery SMC rest	28.5
CD45RO CD4 lymphocyte act	0.3	Coronary artery SMC TNFalpha + IL-1 beta	19.9
CD8 lymphocyte act	0.1	Astrocytes rest	61.6
Secondary CD8 lymphocyte rest	0.2	Astrocytes TNFalpha + IL-1 beta	100.0
Secondary CD8 lymphocyte act	0.6	KU-812 (Basophil) rest	0.3
CD4 lymphocyte none	0.3	KU-812 (Basophil)	0.3

		PMA/ionomycin	
2ry Th1/Th2/Tr1 _anti-CD95 CH11	0.5	CCD1106 (Keratinocytes) none	0.6
LAK cells rest	0.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.9
LAK cells IL-2	0.4	Liver cirrhosis	0.4
LAK cells IL-2+IL-12	0.0	NCI-H292 none	9.5
LAK cells IL-2+IFN gamma	0.6	NCI-H292 IL-4	5.5
LAK cells IL-2+ IL-18	0.2	NCI-H292 IL-9	4.2
LAK cells PMA/ionomycin	0.3	NCI-H292 IL-13	2.5
NK Cells IL-2 rest	0.6	NCI-H292 IFN gamma	1.2
Two Way MLR 3 day	1.4	HPAEC none	0.1
Two Way MLR 5 day	0.6	HPAEC TNF alpha + IL-1 beta	2.2
Two Way MLR 7 day	0.4	Lung fibroblast none	75.8
PBMC rest	0.2	Lung fibroblast TNF alpha + IL-1 beta	11.1
PBMC PWM	5.1	Lung fibroblast IL-4	53.6
PBMC PHA-L	8.3	Lung fibroblast IL-9	27.2
Ramos (B cell) none	0.3	Lung fibroblast IL-13	34.2
Ramos (B cell) ionomycin	0.8	Lung fibroblast IFN gamma	20.4
B lymphocytes PWM	0.4	Dermal fibroblast CCD1070 rest	99.3
B lymphocytes CD40L and IL-4	0.9	Dermal fibroblast CCD1070 TNF alpha	64.6
EOL-1 dbcAMP	0.1	Dermal fibroblast CCD1070 IL-1 beta	64.2
EOL-1 dbcAMP PMA/ionomycin	0.6	Dermal fibroblast IFN gamma	3.3
Dendritic cells none	0.3	Dermal fibroblast IL-4	1.4
Dendritic cells LPS	0.2	Dermal Fibroblasts rest	66.9
Dendritic cells anti-CD40	0.8	Neutrophils TNFa+LPS	0.1
Monocytes rest	0.9	Neutrophils rest	0.0
Monocytes LPS	40.6	Colon	0.1
Macrophages rest	0.1	Lung	8.8
Macrophages LPS	0.5	Thymus	1.2
HUVEC none	0.4	Kidney	0.3
HUVEC starved	0.1		

**AI\_comprehensive panel\_v1.0 Summary:** Ag3626 The NOV42a transcript is expressed in both normal and disease tissue. Transcript expression is higher in some joint tissues isolated from osteoarthritic (OA) patients as compared to normal joint tissues, with highest expression in an OA bone sample (CT=28.5). These findings suggest that the transcript or the protein it encodes could be used to detect osteoarthritic tissues. Furthermore, therapies designed with the protein encoded for by this transcript could be important for the treatment of arthritis.

**CNS\_neurodegeneration\_v1.0 Summary:** Ag3626 This panel does not show differential expression of the NOV42a gene in Alzheimer's disease. However, this expression profile shows the presence of this gene in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic diseases.

5       **General\_screening\_panel\_v1.4 Summary:** Ag3626 Results from one experiment with the NOV42c gene are not included. The amp plot indicates that there were experimental difficulties with this run.

10       **Panel 2.2 Summary:** Ag3626 The expression of this gene appears to be highest in a sample derived from a normal lung tissue (CT=30.8). In addition, there appears to be  
15       substantial expression in other samples derived from lung cancers, bladder cancers, breast cancers, ovarian cancers and colon cancers. Thus, the expression of this gene could be used to distinguish normal lung tissue from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies could be of benefit in the treatment of lung, bladder, breast, ovarian and colon  
15       cancer.

20       **Panel 3D Summary:** Ag3626 The expression of this gene appears to be highest in a sample derived from a lung cancer cell line (LX-1) (CT=27.5). In addition, there appears to be substantial expression in other samples derived from colon cancer cell lines and gastric cancer cell lines. Thus, the expression of this gene could be used to distinguish LX-1 cells from other  
20       samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies could be of benefit in the treatment of colon or gastric cancer.

25       **Panel 4.1D Summary:** Ag3626 Highest expression of the NOV42c gene is seen in TNF-alpha and IL-1 beta treated astrocytes. This expression suggests that therapeutics  
25       designed against the protein encoded by this gene may be useful for the treatment of inflammatory CNS diseases such as multiple sclerosis. In addition, this gene is expressed in clusters of samples from both treated and untreated lung and dermal fibroblasts. Therefore, modulation of the expression or activity of the protein encoded by this transcript may be beneficial for the treatment of lung inflammatory diseases such as asthma, and chronic  
30       obstructive pulmonary diseases, inflammatory skin diseases such as psoriasis, atopic dermatitis, ulcerative dermatitis, ulcerative colitis.

## OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be

5 made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims. The claims presented are representative of the

10 inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

We claim:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
  - a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86;
  - b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;
  - c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86;
  - d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and
  - e) a fragment of any of a) through d).
2. The polypeptide of claim 1 that is a naturally occurring allelic variant of the sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86.
3. The polypeptide of claim 2, wherein the allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n, wherein n is an integer between 1 and 86.
4. The polypeptide of claim 1 that is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution.



5. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically acceptable carrier.
6. A kit comprising in one or more containers, the pharmaceutical composition of claim 5.
7. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic is the polypeptide of claim 1.
8. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:
  - (a) providing the sample;
  - (b) introducing the sample to an antibody that binds immunospecifically to the polypeptide; and
  - (c) determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.
9. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:
  - a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
  - b) comparing the amount of the polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease,wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.
10. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:
  - (a) introducing the polypeptide to the agent; and

- (b) determining whether the agent binds to the polypeptide.
11. The method of claim 10 wherein the agent is a cellular receptor or a downstream effector.
12. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:
- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
  - (b) contacting the cell with a composition comprising a candidate substance; and
  - (c) determining whether the substance alters the property or function ascribable to the polypeptide;
- whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.
13. A method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide of claim 1, the method comprising:
- a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein the test animal recombinantly expresses the polypeptide of claim 1;
  - b) measuring the activity of the polypeptide in the test animal after administering the compound of step (a); and
  - c) comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of claim 1.
14. The method of claim 13, wherein the test animal is a recombinant test animal that expresses a test protein transgene or expresses the transgene under the control of a promoter at an

increased level relative to a wild-type test animal, and wherein the promoter is not the native gene promoter of the transgene.

15. A method for modulating the activity of the polypeptide of claim 1, the method comprising introducing a cell sample expressing the polypeptide of the claim with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

16. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

17. The method of claim 16, wherein the subject is a human.

18. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, or a biologically active fragment thereof.

19. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

- a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 86;
- b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;

- c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86;
  - d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed;
  - e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 86, or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and
  - f) the complement of any of the nucleic acid molecules.
20. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.
21. The nucleic acid molecule of claim 19 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.
22. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 86.
23. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of
- a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86;
  - b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86, is changed from that selected from the group consisting of the

chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed;

- c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86; and
- d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86, is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

24. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 86, or a complement of the nucleotide sequence.

25. The nucleic acid molecule of claim 19, wherein the nucleic acid molecule comprises a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

26. A vector comprising the nucleic acid molecule of claim 19.

27. The vector of claim 26, further comprising a promoter operably linked to the nucleic acid molecule.

28. A cell comprising the vector of claim 27.

29. A method for determining the presence or amount of the nucleic acid molecule of claim 19 in a sample, the method comprising:

- (a) providing the sample;

- (b) introducing the sample to a probe that binds to the nucleic acid molecule; and
  - (c) determining the presence or amount of the probe bound to the nucleic acid molecule,
- thereby determining the presence or amount of the nucleic acid molecule in the sample.

30. The method of claim 29 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

31. The method of claim 30 wherein the cell or tissue type is cancerous.

32. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 19 in a first mammalian subject, the method comprising:

- a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and
- b) comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number  
WO 02/079398 A3

(51) International Patent Classification <sup>7</sup> :	C12N 15/63, 15/12, A61K 38/16, C07K 14/00	60/314,018	21 August 2001 (21.08.2001)	US
		60/315,227	27 August 2001 (27.08.2001)	US
		60/318,403	10 September 2001 (10.09.2001)	US
(21) International Application Number:	PCT/US02/07355	60/318,510	10 September 2001 (10.09.2001)	US
		60/322,360	14 September 2001 (14.09.2001)	US
(22) International Filing Date:	8 March 2002 (08.03.2002)	60/322,296	14 September 2001 (14.09.2001)	US
		60/325,378	27 September 2001 (27.09.2001)	US
(25) Filing Language:	English	60/332,486	9 November 2001 (09.11.2001)	US
		60/345,399	9 November 2001 (09.11.2001)	US
(26) Publication Language:	English	10/094,886	7 March 2002 (07.03.2002)	US

(30) Priority Data:

60/274,194	8 March 2001 (08.03.2001)	US
60/274,281	8 March 2001 (08.03.2001)	US
60/274,322	8 March 2001 (08.03.2001)	US
60/274,849	9 March 2001 (09.03.2001)	US
60/275,578	13 March 2001 (13.03.2001)	US
60/275,579	13 March 2001 (13.03.2001)	US
60/275,601	13 March 2001 (13.03.2001)	US
60/276,000	14 March 2001 (14.03.2001)	US
60/276,776	16 March 2001 (16.03.2001)	US
60/276,994	19 March 2001 (19.03.2001)	US
60/277,239	20 March 2001 (20.03.2001)	US
60/277,327	20 March 2001 (20.03.2001)	US
60/277,338	20 March 2001 (20.03.2001)	US
60/277,791	21 March 2001 (21.03.2001)	US
60/277,833	22 March 2001 (22.03.2001)	US
60/278,152	23 March 2001 (23.03.2001)	US
60/278,894	26 March 2001 (26.03.2001)	US
60/278,999	27 March 2001 (27.03.2001)	US
60/279,036	27 March 2001 (27.03.2001)	US
60/280,233	30 March 2001 (30.03.2001)	US
60/280,802	2 April 2001 (02.04.2001)	US
60/288,052	2 May 2001 (02.05.2001)	US
60/288,066	2 May 2001 (02.05.2001)	US
60/288,228	2 May 2001 (02.05.2001)	US
60/291,766	17 May 2001 (17.05.2001)	US
60/296,693	7 June 2001 (07.06.2001)	US
60/296,856	8 June 2001 (08.06.2001)	US
60/303,230	5 July 2001 (05.07.2001)	US
60/303,237	5 July 2001 (05.07.2001)	US
60/310,913	8 August 2001 (08.08.2001)	US
60/311,978	13 August 2001 (13.08.2001)	US
60/312,191	14 August 2001 (14.08.2001)	US
60/312,916	16 August 2001 (16.08.2001)	US
60/313,182	17 August 2001 (17.08.2001)	US
60/313,626	20 August 2001 (20.08.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	10/094,886 (CIP)
Filed on	7 March 2002 (07.03.2002)
US	60/274,194 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,281 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,322 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/274,849 (CIP)
Filed on	9 March 2001 (09.03.2001)
US	60/275,235 (CIP)
Filed on	12 March 2001 (12.03.2001)
US	60/275,578 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,579 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/275,601 (CIP)
Filed on	13 March 2001 (13.03.2001)
US	60/276,000 (CIP)
Filed on	14 March 2001 (14.03.2001)
US	60/276,776 (CIP)
Filed on	16 March 2001 (16.03.2001)
US	60/276,994 (CIP)
Filed on	19 March 2001 (19.03.2001)
US	60/277,239 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,327 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,338 (CIP)
Filed on	20 March 2001 (20.03.2001)
US	60/277,791 (CIP)
Filed on	21 March 2001 (21.03.2001)
US	60/277,833 (CIP)
Filed on	22 March 2001 (22.03.2001)

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.



WO 02/079398 A3



- US 60/278,152 (CIP)  
 Filed on 23 March 2001 (23.03.2001)  
 US 60/278,894 (CIP)  
 Filed on 26 March 2001 (26.03.2001)  
 US 60/278,999 (CIP)  
 Filed on 27 March 2001 (27.03.2001)  
 US 60/279,036 (CIP)  
 Filed on 27 March 2001 (27.03.2001)  
 US 60/280,233 (CIP)  
 Filed on 30 March 2001 (30.03.2001)  
 US 60/280,802 (CIP)  
 Filed on 2 April 2001 (02.04.2001)  
 US 60/288,052 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/288,066 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/288,228 (CIP)  
 Filed on 2 May 2001 (02.05.2001)  
 US 60/291,766 (CIP)  
 Filed on 17 May 2001 (17.05.2001)  
 US 60/296,693 (CIP)  
 Filed on 7 June 2001 (07.06.2001)  
 US 60/296,856 (CIP)  
 Filed on 8 June 2001 (08.06.2001)  
 US 60/303,230 (CIP)  
 Filed on 5 July 2001 (05.07.2001)  
 US 60/303,237 (CIP)  
 Filed on 5 July 2001 (05.07.2001)  
 US 60/310,913 (CIP)  
 Filed on 8 August 2001 (08.08.2001)  
 US 60/311,978 (CIP)  
 Filed on 13 August 2001 (13.08.2001)  
 US 60/312,191 (CIP)  
 Filed on 14 August 2001 (14.08.2001)  
 US 60/312,916 (CIP)  
 Filed on 16 August 2001 (16.08.2001)  
 US 60/313,182 (CIP)  
 Filed on 17 August 2001 (17.08.2001)  
 US 60/313,626 (CIP)  
 Filed on 20 August 2001 (20.08.2001)  
 US 60/314,018 (CIP)  
 Filed on 21 August 2001 (21.08.2001)  
 US 60/315,227 (CIP)  
 Filed on 27 August 2001 (27.08.2001)  
 US 60/318,403 (CIP)  
 Filed on 10 September 2001 (10.09.2001)  
 US 60/318,510 (CIP)  
 Filed on 10 September 2001 (10.09.2001)  
 US 60/322,296 (CIP)  
 Filed on 14 September 2001 (14.09.2001)  
 US 60/322,360 (CIP)  
 Filed on 14 September 2001 (14.09.2001)  
 US 60/325,378 (CIP)  
 Filed on 27 September 2001 (27.09.2001)  
 US 60/332,486 (CIP)  
 Filed on 9 November 2001 (09.11.2001)  
 US 60/345,399 (CIP)  
 Filed on 9 November 2001 (09.11.2001)
- (71) **Applicant (for all designated States except US):** CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** KEKUDA, Ramesh [IN/US]; 1213 Avalon Valley Drive, Danbury, CT 06810 (US). TCHERNEV, Velizar, T. [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US). LIU, Xiaohong [CN/US]; 96 Montoya Circle, Branford, CT 06405 (US). SPYTEK, Kimberly, A. [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). PATTURAJAN, Meera [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). BURGESS, Catherine, E. [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). VERNET, Corine, A., M. [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). LI, Li [CN/US]; 56 Jerimoth Road, Branford, CT 06405 (US). GORMAN, Linda [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). MALYANKAR, Uriel, M. [IN/US]; 229 Branford Road, # 330, Branford, CT 06405 (US). BOLDOG, Ferenc, L. [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). GUO, Xiaojia [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). SHENOY, Suresh, G. [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). PADIGARU, Muralidhara [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). TAUPIER, Raymond, J., Jr. [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). MILLER, Charles, E. [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). CASMAN, Stacie, J. [US/US]; 17 Peck Street, North Haven, CT 06473 (US). PENA, Carol, E., A. [US/US]; 604 Orange Street, #2, New Haven, CT 06511 (US). GANGOLLI, Esha, A. [IN/US]; 31 Strawberry Hill Road, Madison, CT 06443 (US). GUSEV, Vladimir [UA/US]; 1209 Durham Road, Madison, CT 06443 (US). SMITHSON, Glennnda [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). ZERHUSEN, Bryan, D. [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). GERLACH, Valerie [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). POCHART, Pascale, F.-J. [FR/US]; 99 Country Way, Madison, CT 06443 (US). FERNANDES, Elma, R. [IN/US]; 77 Florence Road #2B, Branford, CT 06405 (US). SHIMKETS, Richard, A. [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). RASTELLI, Luca [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). SPADERNA, Steven, K. [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). LAROCHELLE, William, J. [US/US]; 15 Devonshire Lane, Madison, CT 06443 (US). ZHONG, Mei [CA/US]; Apartment 1B, 45 Harrison Avenue, Branford, CT 06405 (US). KHRAMTSOV, Nikolai, V. [RU/US]; 38 Montoya Circle, Branford, CT 06405 (US). VOSS, Edward, Z. [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). HERRMANN, John, L. [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US).
- (74) **Agent:** ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky, and Popco, P., C., One Financial Center, Boston, MA 02111 (US).
- (81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

[Continued on next page]





MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.

**Published:**

— with international search report

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(88) **Date of publication of the international search report:**

21 August 2003

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07355

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/63, 15/12; A61K 38/16; C07K 14/00

US CL : 435/320.1; 536/23.1; 514/2; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1; 536/23.1; 514/2; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Continuation Sheet

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database SPTREMBL, PETRENKO et al., Accession No. Q63366, Posted 01 November 1997.	1, 2, 5, 6
X	Database Swissprot, PETRENKO et al., Accession No. Q61200, Posted 01 November 1997.	1, 2, 5, 6
X	Database GenBank, GRAVES et al., Accession No. AC004613, Posted 21 December 1999.	19-21, 23-25
—		-----
Y		26-28
Y	WATSON et al., Recombinant DNA Second Edition. New York: W. H. Freeman and Co.. 1994, pages 63-77, especially pages 71-76.	26-28

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"I"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

03 March 2003 (03.03.2003)

Date of mailing of the international search report

27 MAR 2003

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

John S. Brusca

Telephone No. 703 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07355

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-7, 16-18, and SEQ ID NOS: 1 and 2

Remark on Protest

☐  
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

PCT/US02/07355

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: Species 1 and 2 for each of Groups III and VIII are drawn to different methods that comprise different steps and produce different results.

## Continuation of B. FIELDS SEARCHED Item 3:

GenBank sequence database, Geneseq sequence database, issued and published U.S. Patent sequence database, Swissprot sequence database, SPTREMBL sequence database

search terms: SEQ ID NOS: 1 and 2

## INTERNATIONAL SEARCH REPORT

PCT/US02/07355

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-7, and 16-28, drawn to polynucleotides, polypeptides, vectors and host cells comprising the polynucleotides, and a therapeutic method of using the polypeptide (1<sup>st</sup> product and 1<sup>st</sup> method of use)..

Group II, claim(s) 8, drawn to a method of using a polypeptide-specific antibody (2<sup>nd</sup> method).

Group III, claim(s) 9, drawn to a diagnostic or prognostic method comprising an assay of a polypeptide (3<sup>rd</sup> method).

Group IV, claim(s) 10 and 11, drawn to an assay for a ligand that binds a polypeptide (4<sup>th</sup> method).

Group V claim(s) 12-14, drawn to an assay of modulators of activity of a polypeptide.

Group VI, claim(s) 15, drawn to a method of modulating activity of a polypeptide (6<sup>th</sup> method).

Group VII claim(s) 29-31, drawn to a method of assay of a polynucleotide (7<sup>th</sup> method).

Group VIII, claim(s) 32, drawn to a diagnostic or prognostic method comprising an assay of a polynucleotide (8<sup>th</sup> method).

In addition, each Group detailed above reads on distinct Groups drawn to multiple sequences. The sequences are distinct because they are unrelated sequences, and a further lack of unity is applied to each Group. The Applicants must further elect one pair of SEQ ID NOS for a polynucleotide and its encoded polypeptide for examination in the elected Group detailed above. Payment of fees for an additional invention will entitle the Applicants to examination of one additional pair of sequences.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

For Group III, the species are: 1) a diagnostic method, and 2) a prognostic method.

For Group VIII, the species are: 1) a diagnostic method, and 2) a prognostic method.

The claims are deemed to correspond to the species listed above in the following manner:

For Group III, species 1 and 2 correspond to claim 9.

For Group VIII, species 1 and 2 correspond to claim 32.

The following claim(s) are generic: no claims are generic, claims 9 and 32 are Markush-type alternative language claims.

The total number of inventions was calculated based on the number of combinations that exist between the SEQ ID numbers and the total number of groups. The formula is recited below:

$$\text{Total Number of Inventions} = ((\text{number of Groups} + (\text{Number of species} - \text{number of Groups})) \times \text{Total SEQ ID NOS})$$

The inventions listed as Groups I-VIII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: PCT Rule 13.1 and Annex B do not provide for unity of invention between two or more different products, methods of making, or methods of use that share a special technical feature.